RECOMBINATION-MEDIATED MECHANISMS

OF POXVIRUS EVOLUTION

by

Kelsey Rogers Cone

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Human Genetics

The University of Utah

August 2017

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Kelsey Rogers Cone**

has been approved by the following supervisory committee members:

Nels C. Elde	, Chair	**2/28/2017** Date Approved
Mark M. Metzstein	, Member	**2/28/2017** Date Approved
Gabrielle Kardon	, Member	**2/28/2017** Date Approved
Vicente Planelles	, Member	**2/28/2017** Date Approved
Wesley I. Sundquist	, Member	**2/28/2017** Date Approved

and by **Lynn B. Jorde**, Chair/Dean of the Department/College/School of **Human Genetics**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Viruses undergo rapid genetic changes in response to selective forces imposed by host immune defenses. All organisms are at risk of virus infections, and the ancient and ongoing coevolution between viruses and their hosts have shaped genome evolution for both biological partners. One family of viruses, the poxviruses, share a long history with humans, and pose a current threat for emerging and re-emerging diseases. Despite the extensive history of mammalian poxvirus infections, little is known about the ways in which poxviruses adapt to their hosts. In this thesis, I use the model poxvirus, vaccinia virus (VACV), in various experimental evolution systems to investigate mechanisms of poxvirus evolution. These studies reveal new mechanisms of poxvirus adaptation involving recombination-mediated gene amplification events that provide multiple benefits to viral fitness. We first identify adaptive copy number variation (CNV) of a VACV gene encoding a weak virulence factor that promotes fitness in response to host innate immune defenses. Additional gene copies lead to both increased protein expression and an increased chance of gaining beneficial point mutations within the expanded gene locus. I then show that the presence of CNV can also promote the accumulation of a single nucleotide variant in a distant gene. These results establish a direct role for recombination in mediating adaptive genetic changes, while simultaneously promoting additional adaptations in the form of intragenic and extragenic point mutations. To more fully understand the effects of recombination on single nucleotide variant dynamics, I

then analyze evolving VACV populations containing a beneficial point mutation within expanded gene arrays. New sequencing technologies permit detailed analyses of these complex genetic regions, and reveal shared features of genome evolution between VACV and a wide array of organisms that highlight the utility of poxviruses as a model system to investigate mechanisms of rapid gene conversion. These studies collectively reveal new facets of virus adaptation, and add substantially to our understanding of poxvirus evolution with potentially broad applications ranging from devising new therapeutic strategies to predicting and combating future pandemics.

Whilst this planet has gone cycling on according to the fixed law
of gravity, from so simple a beginning endless forms most
beautiful and most wonderful have been,
and are being, evolved.

- Charles Darwin, *The Origin of Species*, 1859

TABLE OF CONTENTS

LIST OF FIGURES

Figures

LIST OF TABLES

Tables

ACKNOWLEDGEMENTS

I would like to thank my mentor and thesis advisor, Nels Elde, for bringing me into his lab. Nels created a lab with excellent science, inspiring questions, and a supportive, fun group of people of which I've been lucky to be a part. I couldn't ask for a better lab: Nels for sharing his creativity and unwavering optimism, Diane for being the best lab mom and an inspiration to us all, the postdocs Matt, Dustin, Ed, Alesia, and Keir for setting the bar high and pushing me to be a better scientist, fellow grad students Della and Clay for the comradery and good times in lab and at retreats, and Missy for making sure lab was never boring. You all made coming to lab worthwhile and enjoyable, and helped keep me sane, motivated, caffeinated, and fed when I needed it, so thank you.

At the University of Utah, I was lucky to have a scientific support system beyond my lab, and even beyond my department. Thank you to all my scientific mentors, Mark Metzstein, Wes Sundquist, Vicente Planelles, and Gab Kardon, for making my dissertation a reality and a success; to Vicente, Matt Mulvey, and Janis Weis for their constant support and mentoring; to all the PIs in Human Genetics for their commitment to facilitate grad student development; and to the Pathology Department and the MPTG group for allowing me to be a part of their communities. A big thank you to all the admins who keep us grad students in line and on track, including Natalie, Kandace, and Stefanie.

I wouldn't have considered a career in science or made it to grad school without

the support of amazing teachers and mentors throughout my life. Thank you to my teachers from grade school through high school: Mr. Hatchen, Dr. Rogers, Mr. Campbell, Mr. McKellips, and Dr. Schoonmaker. Thank you to inspiring scientists at Lewis & Clark College: Dr. Broide, Dr. Lochner, and especially Dr. Lycan. Deborah, thank you for taking a chance and letting me into your upper level class early, and allowing me to work in your lab. Having such a strong female mentor during my first lab experience allowed me to never question my place in science. A big thank you to Dr. Wong for bringing me into your lab, where my love of everything virus really took off. Also thanks to all the Wong Lab members, especially Dr. Estep for teaching me everything I know about virology, and for also teaching me how to keep a good work—life balance.

My life is built on amazing friendships, and I am grateful to have so many true friends from every stage of my life. Sally, Charisse, Apple, Safran, Spector, and Allison, I don't know what I would do without you and all the girls' nights, dog walks, and adventures. Emma, my forever hetero-lifemate, Andreen, Becka, Nate, Roxanne, Garrett, and Alex, college and Portland were everything with you all. Thanks for all the late nights, long talks, and endless fun. To the CO girls, Kip, Kaylie, Erika, Heather, Drea, Sage, and LeAnne, we've been through so much together and I couldn't imagine life without you lovelies. Thank you for your endless love, weirdness, and nose pokes.

Huge thank you to all of my family; I wouldn't be the person I am today without all of your support and love. Thank you to my grandparents, especially Papa, all my aunties, aunts, and uncles, and Alec. My family really is the best. Cones, Sanders, and Percys, thank you for adding me into your families. Sally, Patsy, Craig, Courtney, and Elliott, it's a privilege to be one of the Cones and share in your adventures. Thank you to

my brothers Justin and Conner for being you — nerdy, caring, genuine, fun; I'm lucky to count my siblings as my friends. Thank you to my parents for literally everything. You inspired and encouraged my passions for science, the outdoors, and adventure, and have always been there for me. Finally, and most importantly, thank you to my husband Brian for the unwavering support. You inspire me to be the best I can be, and I can't wait for the adventures to come.

CHAPTER 1

INTRODUCTION

Host-pathogen evolution

Evolutionary conflicts drive genetic change

Host-pathogen interactions are a major driving force in the evolution of every organism. The continuous conflict between pathogen replication and host survival places recurrent selective pressure on each side to adapt. These ongoing bouts of selection are often described as an arms race in which both host and pathogen undergo repeated genetic changes in the battle for evolutionary success. This idea is captured by the classic Red Queen hypothesis first put forth by the evolutionary biologist Leigh Van Valen, in which he borrows from Lewis Carroll's book *Through the Looking Glass*, when the Red Queen tells Alice that "it takes all the running you can do, to stay in the same place" (Van Valen, 1973). Indeed, despite billions of years of evolution, host-pathogen interactions continue to drive adaptation in an endless contest for survival.

At the molecular level, adaptations often occur at protein-protein interfaces where host and pathogen-encoded factors directly interact. These interactions arise because pathogens utilize a range of host factors to promote replication, and hosts deploy numerous factors to detect and defend against infection. Thus, different types of interactions dictate the kind of adaptation observed at the protein level. From the host

perspective, proteins hijacked or antagonized by pathogens are often evolving defensively to alter surfaces such that the interacting pathogen protein can no longer bind (Daugherty and Malik, 2012). Alternatively, host defense proteins targeting specific pathogen functions evolve offensively to maintain the interaction (Daugherty and Malik, 2012). Changes in host proteins place selective pressure on the pathogen to evolve compensating mutations to reestablish or abolish the interaction. These back-and-forth adaptations on specific protein surfaces create hot spots of rapid evolution for both hosts and pathogens, which can be detected in genomic sequences.

Within genomes, the random occurrence of point mutations generates variation upon which selection can act. In protein-coding genes, synonymous mutations, or nucleotide changes that do not alter the amino acid sequence, are often neutral and have a stochastic chance of becoming fixed in a population over time. Nonsynonymous mutations, which do change the resulting amino acid, can have varying fitness effects with diverse outcomes in response to selection. Deleterious mutations affecting protein function are selected against, implying constraint. Over evolutionary time, constraint is detected in the form of amino acid conservation, also known as purifying selection (Daugherty and Malik, 2012). Highly variable residues instead suggest that nonsynonymous mutations are being selected for, and these are examples of positive or diversifying selection (Holmes, 2004; Hurst, 2002). Comparing the rate of nonsynonymous to synonymous point mutations (dN/dS) within a gene across large phylogenetic distances provides a means of detecting instances of purifying ($dN/dS < 1$) or positive ($dN/dS > 1$) selection (Hurst, 2002; Pond et al., 2005; Yang, 2007). Genes encoding proteins involved in molecular arms races are often under positive selection,

due to recurrent selective pressure acting to fix nonsynonymous changes that alter binding interfaces (Daugherty and Malik, 2012). Evolutionary analyses aid the identification of Red Queen conflicts between pathogen and host proteins (Daugherty and Malik, 2012; Holmes, 2004), and studying these arms races can help to understand how both hosts and pathogens evolve.

Evolution of host genomes in response to pathogens

Complex organisms, such as mammals, must compete with pathogens which possess higher mutation rates, larger population sizes, and shorter reproductive times. The odds seem stacked against the host, yet mammalian genomes have the advantage of size, and are capable of encoding a great variety of genes dedicated to defending against pathogens. Complexity and diversification of immune genes in response to a long history of infections has shaped mammalian immune responses to greatly constrain the evolutionary success of pathogens (Daugherty and Malik, 2012; Meyerson and Sawyer, 2011; tenOever, 2016). The large genome size of mammals is also in part a direct consequence of infections, through the incorporation of retroviruses into the genome of their host. Endogenized retroviruses make up a large portion of many animal genomes with a significant impact on various processes, emphasizing the importance of the ancient relationship between viruses and their hosts (Bannert and Kurth, 2006; Katzourakis et al., 2005). Analyses of host genomes also provide evidence of various adaptive outcomes in response to infection. While many adaptations reveal a history of selection for changes beneficial to the host, there are cases where adaptation to avoid infection may not increase host fitness overall. Antagonistic pleiotropy can arise when evolutionary

tradeoffs occur such that mutations beneficial in one environment are detrimental in another. One of the most well-known examples is the maintenance of hemoglobin alleles in humans that provide resistance to malaria, at the expense of sickle cell anemia in homozygous individuals (Ashley-Koch et al., 2000). These different outcomes demonstrate how varied adaptive strategies play out over time, resulting in fixed changes in host genomes that belie the volatile history of interactions with various pathogenic threats. Evolutionary analyses can uncover these adaptations, which aid in understanding the effects driving changes in various genomes and animal populations.

Identification and analysis of molecular arms races demonstrate the power of using an evolutionary perspective to reveal how genetic changes can rapidly influence host adaptation. Phylogenetic analyses across different primate species have shown that genes under positive selection are enriched for those encoding immunity factors (George et al., 2011; Nielsen et al., 2005). These include a number of well-characterized pattern recognition receptors and restriction factors, which are host-encoded proteins that detect and inhibit infections, respectively. The observed signatures of diversifying selection are likely a result of repeated antagonism by pathogens such as viruses (Daugherty and Malik, 2012; Duggal and Emerman, 2012) and bacteria (Wlasiuk and Nachman, 2010) during Red Queen conflicts. Molecular arms races can also involve host proteins required for essential cellular functions, highlighting the complex interactions between pathogens and their hosts. For example, the requirement for transition metals has resulted in the regulation and sequestration of metals by hosts, termed “nutritional immunity”, and scavenging of metals by pathogenic bacteria (Hood and Skaar, 2012) and fungi (Potrykus et al., 2014). Transition metals are required in many essential biological processes, and

thus host transport proteins may be subject to evolutionary constraint. However, recently, it was shown that the iron transport protein transferrin in primates is involved in an ongoing conflict with the bacterial iron scavenging protein TbpA (Barber and Elde, 2014). In this study, signatures of positive selection in transferrin helped identify critical residues at the TbpA binding interface, and validated functional consequences of genetic variation in transferrin for resistance to bacterial pathogens. This type of detailed evolutionary analysis demonstrates the ability to uncover new aspects of adaptation in response to pathogens, and has been applied to a number of host-pathogen interactions.

The collective study of molecular arms races identified using markers of positive selection reveals patterns that are beginning to guide our understanding of how these interactions shape evolutionary change (Daugherty and Malik, 2012). Virus-host interactions in mammals provide a particularly useful system to understand adaptive strategies relevant to human health and evolution. For example, specific residues under positive selection can alter interactions between host and pathogen proteins, which was first demonstrated by examining the interaction between the antiviral factor tripartite motif-containing protein 5 alpha (TRIM5α) and human immunodeficiency virus 1 (HIV-1) (Sawyer et al., 2005). In this study, the authors showed that swapping just five amino acids under positive selection from the restrictive rhesus monkey protein into human TRIM5α conferred resistance to HIV-1. It was later shown that a single amino acid could largely determine the specificity of the interaction between TRIM5α and the HIV-1 capsid (Stremlau et al., 2005; Yap et al., 2005), and a similar observation was made for the interaction between the restriction factor SAM and HD domain containing deoxynucleoside triphosphate triphosphohydrolase 1 (SAMHD1) and the lentivirus

protein Vpx (Laguette et al., 2012; Lim et al., 2012). Thus, even single amino acid changes can have a dramatic impact on the ability of host restriction factors to bind viral products, which in turn can determine whether a host is susceptible to infection. Taking an evolutionary approach to study this type of interaction can therefore elucidate critical residues involved in immune defenses, providing new information to guide therapeutic approaches. Additionally, these analyses uncover the various ways in which host genomes have been shaped by infections, with the potential to inform our understanding of future infection dynamics.

Evolution of pathogens in response to host defenses

With the continuous threat of existing and emerging infections, it is essential to understand how pathogens evolve in response to immune responses and medical treatments. Pathogens are capable of extremely rapid change, dictated largely by their interactions with a host. By definition, a pathogen must interact with a host, and consequently face intense selective pressure from host defenses. Pathogens utilize a variety of adaptive strategies in response, including high point mutation rates relative to their hosts for the rapid sampling of variants (Drake et al., 1998; Lynch, 2010; Sniegowski et al., 2000). Particularly for bacteria, horizontal gene transfer, or the exchange of genetic material, is another critical source of adaptation (reviewed in de la Casa-Esperón, 2012; Koonin, 2016). Acquisition of an entire gene can deliver immediate innovation, and has also been observed in some large DNA viruses (Filée et al., 2008; Hughes and Friedman, 2005). Amplification of existing genes in bacteria similarly generates extensive genetic variation on which evolution can act. Gene amplification

events in response to host pressures can promote enhanced survival in the host environment, and have also been shown to increase virulence (Andersson and Hughes, 2009). These few examples highlight the complex nature of pathogen evolution, and indeed a multitude of adaptive mechanisms are employed by the diverse array of pathogens. However, general principles and strategies can still be discovered when studying individual pathogens as they evolve.

At the molecular level, the evolution of pathogen proteins involved in Red Queen conflicts share many similarities with mechanisms of host adaptation. As in the examples listed above for host genes, positive selection can be used to identify pathogen genes undergoing rapid evolution during ongoing arms race scenarios (reviewed in Aguileta et al., 2009). While analyses of dN/dS ratios can be more complicated in pathogen genomes, due to confounding effects like bottlenecks, horizontal gene transfer, and recombination, they can still aid in identification of host-pathogen arms races. Pathogen-encoded proteins that directly interact with host proteins showing signatures of positive selection are often also rapidly evolving. For example, the bacterial receptor TbpA that binds host transferrin shows evidence of recurrent selection at the interaction interface (Barber and Elde, 2014). Similarly, a common finding among viruses is the concentration of positively selected residues on surface-exposed proteins, suggesting direct interaction with host factors. In some instances, these proteins are likely under selection from recognition by host adaptive immune system components, as in the case of dengue virus surface proteins (Bennett et al., 2006), influenza virus hemagglutinin and neuraminidase (Li et al., 2011), and the SARS coronavirus spike protein (Zhang et al., 2006). Selection to bind host receptors can also elicit changes to surface proteins, as observed in the

glycoprotein of Ebolavirus (de La Vega et al., 2015). Small differences affecting host receptor binding can even lead to changes in host range, as in the well-studied emergence of canine parvovirus in dogs through alterations to the major capsid protein (Hueffer and Parrish, 2003; Shackelton et al., 2005). Thus, analyses of positive selection in pathogen genomes can identify evolutionary hot spots, and reveal important innovations that provide insight into various aspects of pathogen evolution, including pathogenesis, virulence, and tropism.

While comparative analyses of existing organisms provide a means of identifying ongoing host-pathogen arms races, there are still many open questions about the dynamics of evolutionary change. The field of experimental evolution allows for more direct observations of evolution in action and testing hypotheses about the processes involved. Microorganisms are particularly amenable to experimental evolution due to large population sizes, short generation times, and asexual reproduction. This permits observation of genotypic and phenotypic changes in microbial populations under controlled conditions over many generations. Studies using microbes, namely viruses, bacteria, and yeast, have observed how fitness changes over time, and how the rate of adaptation is affected by parameters such as population size, population structure, and environment (reviewed in Barrick and Lenski, 2013; Elena and Lenski, 2003). The advent of new sequencing technologies has also enhanced the ability to track genetic changes, leading to new insights such as the role of epistasis in genome evolution (reviewed in Jerison and Desai, 2015). These findings support broad evolutionary mechanisms shared among different organisms, and highlight the utility of microbial experiments to understand general evolutionary principles.

While the discovery of common adaptive mechanisms reinforces evolutionary theory, it is also important to consider the dynamics of adaptation for specific types of microbes. The huge impact of pathogenic microbes on human and animal health necessitates a better understanding of the specific mechanisms driving pathogen evolution. Viruses provide a particularly useful system to study evolutionary processes, as they are associated with every life form, and thus have influenced host-pathogen evolution since the beginning of cellular life (Koonin and Dolja, 2013). Experimental evolution studies using phages have been used to advance our understanding of viral-host adaptation, with applications to general virus biology and evolution (Hall et al., 2013). This framework exploring viral evolutionary processes is also being applied to medically relevant viruses, including the impact of codon usage in poliovirus (Lauring et al., 2012), and different selection outcomes for arboviruses in mosquito or vertebrate hosts with West Nile virus (Deardorff et al., 2011; Jerzak et al., 2007; 2008), chikungunya virus (Coffey and Vignuzzi, 2011), and dengue virus (Vasilakis et al., 2009). These studies can have major implications towards our understanding of viral success, demonstrated by the selection and emergence of epidemic chikungunya virus strains in an experimental system (Stapleford et al., 2014). Thus, this type of experimentation can reveal new mechanisms of viral evolution, which in turn could help inform strategies to combat viral infections. Combining evolutionary perspectives from both the virus and host viewpoints exposes mechanisms of adaptation and counter-adaptation, affording a deeper understanding of the forces shaping infection dynamics. Information regarding the molecular determinants of resistance and susceptibility could someday even enable predictions about pathogen evolution to better support infection control.

Poxviruses

Introduction to poxviruses

The virus family *Poxviridae* is comprised of large, complex viruses containing linear double-stranded DNA (dsDNA) genomes. These viruses replicate solely in the cytoplasm of host cells, and infect a wide range of arthropod (Entomopoxviruses) and vertebrate (Chordopoxviruses) hosts (Moss, 2013a). Among the poxviruses are medically important species capable of infecting humans, most notably variola virus, the causative agent of smallpox. Indeed, humans have a long history with poxvirus infections (Behbehani, 1983), and some estimates even suggest that smallpox has killed more humans in recorded history than all other infections combined (McFadden, 2005). While the successful eradication of smallpox through worldwide vaccination programs has eliminated this threat, other poxviruses capable of zoonotic infections, including monkeypox and cowpox, pose a risk to human health through outbreaks or potential acts of bioterrorism (Essbauer et al., 2010). Emerging and re-emerging viral diseases represent an area of particular concern, due to a lack of treatments and the lag time in creating vaccines in response to epidemics (McCloskey et al., 2014; Morens et al., 2004). It is therefore critical to understand zoonotic threats, and to do this, more information is needed regarding how different types of viruses function and evolve.

Most studies of poxviruses concentrate on the Orthopoxvirus genus of the Chordopoxviruses, since this clade includes variola virus and others with zoonotic potential. Importantly, the Orthopoxvirus genus also includes vaccinia virus (VACV), which was the vaccine strain used to eradicate smallpox. VACV is also the model poxvirus, and most of the current knowledge regarding poxvirus biology and immune

modulation stems from studying vaccinia. The VACV genome is representative of poxvirus complexity, in terms of both structure and composition. About 200 genes are encoded within the ~200kbp VACV genome (Goebel et al., 1990), approximately half of which are shared among Chordopoxviruses, and 49 of which are shared among all poxviruses (Upton et al., 2003). These conserved poxvirus genes are concentrated in the center of the genome, and encode proteins with essential functions in processes such as entry, replication, transcription, and virion assembly (Gubser et al., 2004; Upton et al., 2003). Genes on either end of the genome are more variable between poxvirus species, and encode many virulence genes that modulate host immune responses (Gubser et al., 2004; Moss and Shisler, 2001; Seet et al., 2003). Analyses of poxvirus phylogeny and evolution (reviewed in Lefkowitz et al., 2006) show a greater similarity between poxvirus proteins and eukaryotic rather than prokaryotic proteins, supporting recent and ancient gene acquisition from hosts via horizontal gene transfer (Bratke and McLysaght, 2008; Hughes and Friedman, 2005; McLysaght et al., 2003; Odom et al., 2009). Poxvirus genome size and complexity both contribute to the evolutionary success of these viruses in a wide range of hosts.

The poxvirus replication cycle has been best studied in VACV, and entails binding, entry, transcription of viral genes, replication of the DNA genome, and assembly and morphogenesis of new viral particles (Figure 1.1). Unlike many other viruses, poxvirus replication takes place solely within the host cell cytoplasm, and new virions are generated after about 8 hours (Moss, 2013a). Another unique feature of poxviruses is the existence of two main forms of infectious particles, the intracellular mature virion (IMV) and the extracellular enveloped virion (EEV), thought to facilitate infection between

hosts or within a host, respectively (Moss, 2013a). Both the IMV and EEV forms can enter cells at the plasma membrane or following marcopinocytosis, using at least 15 viral proteins for attachment and membrane fusion (reviewed in Moss, 2012; 2016; Schmidt et al., 2012). After virion fusion, early transcription takes place within the viral core, in which the viral RNA polymerase and encapsidated viral transcription factors transcribe the viral set of early mRNAs (Broyles, 2003; Moss, 2013a). Next, the viral DNA is released into the cytoplasm following core dissolution, where it functions as a template for genome replication (Moss 2013b). DNA replication factors are among those transcribed as early genes, including the viral DNA polymerase, a helicase-primase, a processivity factor, a DNA ligase, and a single-stranded DNA binding protein (reviewed in Moss, 2013b). Post-replicative transcription continues in a temporal manner, in which intermediate gene transcription requires early products, and late gene transcription requires intermediate products (Broyles, 2003; Yang et al., 2013). Intermediate and late gene transcripts encode structural proteins and other factors required for virion assembly and morphogenesis, as well as early transcription factors that are packaged in the virion (Broyles, 2003). Virion morphogenesis takes place in virus factories within the host cell cytoplasm, utilizing host membranes to construct the different virion forms. The IMV form is made first, surrounded by a single membrane, and remains intracellular until cell lysis (Roberts and Smith, 2008; Smith and Law, 2004). Some IMVs are further enveloped by host membrane and transported to the cell surface, where they are released as EEVs (Roberts and Smith, 2008; Smith and Law, 2004). Many of these general features of VACV replication are highly conserved among poxviruses (Moss, 2013a), demonstrating the utility of VACV as a model for studying poxvirus biology.

Orthopoxvirus host range and immune modulation

The host range of different Orthopoxviruses varies widely, from single host specificity (variola virus) to a large variety of hosts (cowpox virus), likely due to the diversity in encoded virulence factors (Haller et al., 2014). At the cellular level, tropism is determined downstream of binding and entry, as no specific host receptors have been identified (McFadden, 2005). The specific interactions between the host cellular immune response and host range factors encoded by the virus thus determine whether a productive infection takes place (Bratke et al., 2013; Haller et al., 2014; Werden et al., 2008). Many of the poxvirus virulence factors affecting host immune responses were first identified through shared homology with cellular gene products (Bugert and Darai, 2000). These represent examples of protein mimicry, in which the viral gene likely arose through acquisition by horizontal gene transfer from a host followed by divergent evolution (Elde and Malik, 2009). Protein mimics and other immune modulatory proteins comprise the set of poxvirus virulence factors capable of influencing various host defense pathways to mount a successful infection.

Poxvirus virulence factors induce changes to the host immune system through three main mechanisms: modulating extracellular communication signals to promote a protected microenvironment, altering intracellular signaling pathways, and masking signs of infection to reduce activation of host defense pathways (reviewed in Moss and Shisler, 2001; Seet et al., 2003). In the extracellular environment, poxviruses target key immune signaling molecules, including interferons, tumor necrosis factor, interleukins, cytokines, chemokines, and complement (Seet et al., 2003). These effects are often elicited by viroceptors, altered and sometimes secreted versions of cellular receptors that compete

for ligands, and virokines, secreted viral proteins that mimic host cytokines or their inhibitors (Johnston and McFadden, 2004; Moss and Shisler, 2001; Seet et al., 2003). The combined action of these viral effectors is to downregulate the early inflammatory response, creating a local environment that favors establishment of infection. Within infected cells, poxviruses also modulate a number of pathways to promote a favorable cellular environment. This includes disruption of apoptotic, interferon receptor, and IL-1β receptor signaling pathways, to downregulate cell-intrinsic innate immune responses (Johnston and McFadden, 2003; Moss and Shisler, 2001; Seet et al., 2003). Poxviruses also encode virulence factors that thwart cell detection pathways, blocking immune responses at the earliest stages of infection. Strategies include targeting host pattern recognition receptors (PRRs) that detect features common among pathogens, as well as masking these pathogen-associated molecular patterns to prevent PRR activation (Johnston and McFadden, 2003; Moss and Shisler, 2001; Seet et al., 2003). Collectively, the numerous immune evasion techniques employed by poxviruses emphasize their complex interactions with a host, and highlight the importance for these viruses to overcome early immune responses.

Specific interfaces between poxviruses and cellular immune defenses illustrate the dynamic nature of host-pathogen interactions. The interface between the host PRR protein kinase R (PKR) and VACV host range factors is an illustrative example of the complex interactions that can determine the outcome of infection (Figure 1.2). PKR is a broadly antiviral cytoplasmic host nucleic acid sensor, activated by double-stranded RNA (dsRNA) produced during many viral infections (Weber et al., 2006). Upon activation, PKR phosphorylates its substrate, eukaryotic translation initiation factor 2α (eIF2α),

leading to a global block in protein translation (Meurs et al., 1990). PKR activation also induces apoptosis (Gil and Esteban, 2000), thus impeding viral replication on multiple fronts. This places selection on viruses to inhibit PKR activation, and indeed a wide variety of viruses encode proteins that antagonize PKR (Langland et al., 2006). In turn, repeated and diverse viral antagonism has resulted in the rapid evolution of PKR in mammals during this complex molecular arms race (Elde et al., 2009; Rothenburg et al., 2009). The interaction in poxviruses is no less complex, with two separate PKR inhibitors encoded by many Orthopoxviruses (Bratke et al., 2013). In VACV, the products of the genes E3L and K3L both inhibit PKR (Beattie et al., 1991; Carroll et al., 1993; Chang et al., 1992), although through different mechanisms (Davies et al., 1993). E3L blocks PKR activation by binding dsRNA (Watson et al., 1991) and thus also inhibits other dsRNA binding proteins such as oligoadenylate synthetase 1 (OAS-1) (Davies et al., 1993) and adenosine deaminase acting on RNA 1 (ADAR1) (Liu et al., 2001). K3L is a protein mimic of eIF2α, and therefore specifically inhibits PKR as a pseudosubstrate (Dar and Sicheri, 2002). Both E3L and K3L are host range factors required for optimal fitness in different host species, which likely contributes to the retention of two independent PKR inhibitors in VACV (Langland and Jacobs, 2002). Examination of host-pathogen interactions in this single example highlights how the interplay between viral and host fitness can result in complicated relationships arising from different evolutionary strategies. These relationships can then be leveraged to test how a virus or host will adapt to selection, providing a more precise comprehension of the adaptive strategies being used.

Conclusions

The ancient and ongoing conflicts between pathogens and their hosts have shaped genome evolution, leaving behind genetic imprints that can be used to trace adaptive strategies and outcomes for diverse organisms. Identification and analyses of genetic conflicts driven by host-pathogen interactions provide a wealth of information to understand the ways in which our own evolutionary history has been shaped by infections, as well as the various tactics pathogens employ to persist and propagate. This knowledge can be used to inform efforts to fight current and future pathogenic threats. Emerging and re-emerging diseases with pandemic potential are a global issue, particularly in the current age of worldwide travel and environmental changes increasing the chance of zoonoses (Morens et al., 2004). Lessons from history aid preparedness, but it is impossible to predict what form the next epidemic will take. Therefore, it is critical to study a diverse array of infectious agents, particularly those with pandemic potential, to be equipped with an understanding of how pathogens function and evolve.

Viruses are an enormously diverse group of obligate intracellular pathogens that have coevolved with all forms of life. The remarkable diversity in size, shape, genome complexity, and gene content among viruses exemplifies the varied and complex nature of virus-host interactions. Numerous viruses pose a serious threat to human and animal health, but to group viral diseases together greatly oversimplifies the incredible variety of virus types and adaptive strategies. Poxviruses are a virus family with a long history of infection in animals. Estimates of smallpox origins make it one of the oldest human pathogens, and it was also the first pathogen to be eradicated through the development and implementation of vaccination (Behbehani, 1983). Despite the history and extensive

study of the related vaccinia virus, little is known about how poxviruses evolve. The potential exists for poxviruses to cause epidemics through zoonosis or bioterrorism (Essbauer et al., 2010), and thus a deeper understanding of the ways in which these viruses adapt is needed. VACV provides a tractable model system in which to study poxvirus evolution, as studies with VACV have uncovered most of what is currently known about poxvirus biology. The work presented in this dissertation utilizes an experimental evolution system to study VACV biology, which reveals new mechanisms of poxvirus adaptation. Application of an evolutionary framework to study virus-host interactions exposes the intricate, complex, and sometimes unexpected nature of selection and adaptation in different systems to inform our understanding of biology and evolutionary success.

References

Aguileta, G., Refrégier, G., Yockteng, R., Fournier, E., and Giraud, T. (2009). Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. Infect. Genet. Evol. *9*, 656–670.

Andersson, D.I., and Hughes, D. (2009). Gene amplification and adaptive evolution in bacteria. Annu. Rev. Genet. *43*, 167–195.

Ashley-Koch, A., Yang, Q., and Olney, R.S. (2000). Sickle hemoglobin (HbS) allele and sickle cell disease: a HuGE review. Am. J. Epidemiol. *151*, 839–845.

Bannert, N., and Kurth, R. (2006). The evolutionary dynamics of human endogenous retroviral families. Annu. Rev. Genomics Hum. Genet. *7*, 149–173.

Barber, M.F., and Elde, N.C. (2014). Nutritional immunity. Escape from bacterial iron piracy through rapid evolution of transferrin. Science *346*, 1362–1366.

Barrick, J.E., and Lenski, R.E. (2013). Genome dynamics during experimental evolution. Nat. Rev. Genet. *14*, 827–839.

Beattie, E., Tartaglia, J., and Paoletti, E. (1991). Vaccinia virus-encoded eIF-2 alpha homolog abrogates the antiviral effect of interferon. Virology *183*, 419–422.

Behbehani, A.M. (1983). The smallpox story: life and death of an old disease. Microbiol. Rev. *47*, 455–509.

Bennett, S.N., Holmes, E.C., Chirivella, M., Rodriguez, D.M., Beltran, M., Vorndam, V., Gubler, D.J., and McMillan, W.O. (2006). Molecular evolution of dengue 2 virus in Puerto Rico: positive selection in the viral envelope accompanies clade reintroduction. J. Gen. Virol. *87*, 885–893.

Bratke, K.A., and McLysaght, A. (2008). Identification of multiple independent horizontal gene transfers into poxviruses using a comparative genomics approach. BMC Evol. Biol. *8*, 67.

Bratke, K.A., McLysaght, A., and Rothenburg, S. (2013). A survey of host range genes in poxvirus genomes. Infect. Genet. Evol. *14*, 406–425.

Broyles, S.S. (2003). Vaccinia virus transcription. J. Gen. Virol. *84*, 2293–2303.

Bugert, J.J., and Darai, G. (2000). Poxvirus homologues of cellular genes. Virus Genes *21*, 111–133.

Carroll, K., Elroy-Stein, O., Moss, B., and Jagus, R. (1993). Recombinant vaccinia virus K3L gene product prevents activation of double-stranded RNA-dependent, initiation factor 2 alpha-specific protein kinase. J. Biol. Chem. *268*, 12837–12842.

Chang, H.W., Watson, J.C., and Jacobs, B.L. (1992). The E3L gene of vaccinia virus encodes an inhibitor of the interferon-induced, double-stranded RNA-dependent protein kinase. Proc. Natl. Acad. Sci. U.S.A. *89*, 4825–4829.

Coffey, L.L., and Vignuzzi, M. (2011). Host alternation of chikungunya virus increases fitness while restricting population diversity and adaptability to novel selective pressures. J. Virol. *85*, 1025–1035.

Dar, A.C., and Sicheri, F. (2002). X-ray crystal structure and functional analysis of vaccinia virus K3L reveals molecular determinants for PKR subversion and substrate recognition. Mol. Cell *10*, 295–305.

Daugherty, M.D., and Malik, H.S. (2012). Rules of engagement: molecular insights from host-virus arms races. Annu. Rev. Genet. *46*, 677–700.

Davies, M.V., Chang, H.W., Jacobs, B.L., and Kaufman, R.J. (1993). The E3L and K3L vaccinia virus gene products stimulate translation through inhibition of the double-stranded RNA-dependent protein kinase by different mechanisms. J. Virol. *67*, 1688–1692.

de la Casa-Esperón, E. (2012). Horizontal transfer and the evolution of host-pathogen interactions. Int. J. Evol. Biol. *2012*, 679045.

de La Vega, M.A., Stein, D., and Kobinger, G.P. (2015). Ebolavirus evolution: past and

present. PLoS Pathog. *11*, e1005221.

Deardorff, E.R., Fitzpatrick, K.A., Jerzak, G.V., Shi, P.Y., Kramer, L.D., and Ebel, G.D. (2011). West Nile virus experimental evolution in vivo and the trade-off hypothesis. PLoS Pathog. *7*, e1002335.

Drake, J.W., Charlesworth, B., Charlesworth, D., and Crow, J.F. (1998). Rates of spontaneous mutation. Genetics *148*, 1667–1686.

Duggal, N.K., and Emerman, M. (2012). Evolutionary conflicts between viruses and restriction factors shape immunity. Nat. Rev. Immunol. *12*, 687–695.

Elde, N.C., and Malik, H.S. (2009). The evolutionary conundrum of pathogen mimicry. Nat. Rev. Microbiol. *7*, 787–797.

Elde, N.C., Child, S.J., Geballe, A.P., and Malik, H.S. (2009). Protein kinase R reveals an evolutionary model for defeating viral mimicry. Nature *457*, 485–489.

Elena, S.F., and Lenski, R.E. (2003). Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat. Rev. Genet. *4*, 457–469.

Essbauer, S., Pfeffer, M., and Meyer, H. (2010). Zoonotic poxviruses. Vet. Microbiol. *140*, 229–236.

Filée, J., Pouget, N., and Chandler, M. (2008). Phylogenetic evidence for extensive lateral acquisition of cellular genes by Nucleocytoplasmic large DNA viruses. BMC Evol. Biol. *8*, 320.

George, R.D., McVicker, G., Diederich, R., Ng, S.B., MacKenzie, A.P., Swanson, W.J., Shendure, J., and Thomas, J.H. (2011). Trans genomic capture and sequencing of primate exomes reveals new targets of positive selection. Genome Res. *21*, 1686–1694.

Gil, J., and Esteban, M. (2000). Induction of apoptosis by the dsRNA-dependent protein kinase (PKR): mechanism of action. Apoptosis *5*, 107–114.

Goebel, S.J., Johnson, G.P., Perkus, M.E., Davis, S.W., Winslow, J.P., and Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. Virology *179*, 247-266, 517-563.

Gubser, C., Hué, S., Kellam, P., and Smith, G.L. (2004). Poxvirus genomes: a phylogenetic analysis. J. Gen. Virol. *85*, 105–117.

Hall, J.P., Harrison, E., and Brockhurst, M.A. (2013). Viral host-adaptation: insights from evolution experiments with phages. Curr. Opin. Virol. 3, 572-577.

Haller, S.L., Peng, C., McFadden, G., and Rothenburg, S. (2014). Poxviruses and the evolution of host range and virulence. Infect. Genet. Evol. *21*, 15–40.

Holmes, E.C. (2004). Adaptation and immunity. PLoS Biol. *2*, e307.

Hood, M.I., and Skaar, E.P. (2012). Nutritional immunity: transition metals at the pathogen–host interface. Nat. Rev. Microbiol. *10*, 525–537.

Hueffer, K., and Parrish, C.R. (2003). Parvovirus host range, cell tropism and evolution. Curr. Opin. Microbiol. *6*, 392–398.

Hughes, A.L., and Friedman, R. (2005). Poxvirus genome evolution by gene gain and loss. Mol. Phylogenet. Evol. *35*, 186–195.

Hurst, L.D. (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. *18*, 486.

Jerison, E.R., and Desai, M.M. (2015). Genomic investigations of evolutionary dynamics and epistasis in microbial evolution experiments. Curr. Opin. Genet. Dev. *35*, 33–39.

Jerzak, G.V., Bernard, K., Kramer, L.D., Shi, P.Y., and Ebel, G.D. (2007). The West Nile virus mutant spectrum is host-dependant and a determinant of mortality in mice. Virology *360*, 469–476.

Jerzak, G.V., Brown, I., Shi, P.Y., Kramer, L.D., and Ebel, G.D. (2008). Genetic diversity and purifying selection in West Nile virus populations are maintained during host switching. Virology *374*, 256–260.

Johnston, J.B., and McFadden, G. (2003). Poxvirus immunomodulatory strategies: current perspectives. J. Virol. *77*, 6093–6100.

Johnston, J.B., and McFadden, G. (2004). Technical knockout: understanding poxvirus pathogenesis by selectively deleting viral immunomodulatory genes. Cell Microbiol. *6*, 695–705.

Katzourakis, A., Rambaut, A., and Pybus, O.G. (2005). The evolutionary dynamics of endogenous retroviruses. Trends Microbiol. *13*, 463–468.

Koonin, E.V. (2016). Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. F1000Res. *5*, 1805–1809.

Koonin, E.V., and Dolja, V.V. (2013). A virocentric perspective on the evolution of life. Curr. Opin. Virol. *3*, 546-557.

Laguette, N., Rahm, N., Sobhian, B., Chable-Bessia, C., Münch, J., Snoeck, J., Sauter, D., Switzer, W.M., Heneine, W., Kirchhoff, F., Delsuc, F., Telenti, A., and Benkirane, M. (2012). Evolutionary and functional analyses of the interaction between the myeloid restriction factor SAMHD1 and the lentiviral Vpx protein. Cell Host Microbe *11*, 205–217.

Langland, J.O., and Jacobs, B.L. (2002). The role of the PKR-inhibitory genes, E3L and K3L, in determining vaccinia virus host range. Virology *299*, 133–141.

Langland, J.O., Cameron, J.M., Heck, M.C., Jancovich, J.K., and Jacobs, B.L. (2006). Inhibition of PKR by RNA and DNA viruses. Virus Res. *119*, 100–110.

Lauring, A.S., Acevedo, A., Cooper, S.B., and Andino, R. (2012). Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. Cell Host Microbe *12*, 623–632.

Lefkowitz, E.J., Wang, C., and Upton, C. (2006). Poxviruses: past, present and future. Virus Res. *117*, 105–118.

Li, W., Shi, W., Qiao, H., Ho, S.Y., Luo, A., Zhang, Y., and Zhu, C. (2011). Positive selection on hemagglutinin and neuraminidase genes of H1N1 influenza viruses. Virology J. *8*, 183.

Lim, E.S., Fregoso, O.I., McCoy, C.O., Matsen, F.A., Malik, H.S., and Emerman, M. (2012). The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. Cell Host Microbe *11*, 194–204.

Liu, Y., Wolff, K.C., Jacobs, B.L., and Samuel, C.E. (2001). Vaccinia virus E3L interferon resistance protein inhibits the interferon-induced adenosine deaminase A-to-I editing activity. Virology *289*, 378–387.

Lynch, M. (2010). Evolution of the mutation rate. Trends Genet. *26*, 345–352.

McCloskey, B., Dar, O., Zumla, A., and Heymann, D.L. (2014). Emerging infectious diseases and pandemic potential: status quo and reducing risk of global spread. Lancet Infect. Dis. *14*, 1001–1010.

McFadden, G. (2005). Poxvirus tropism. Nat. Rev. Microbiol. *3*, 201–213.

McLysaght, A., Baldi, P.F., and Gaut, B.S. (2003). Extensive gene gain associated with adaptive evolution of poxviruses. Proc. Natl. Acad. Sci. U.S.A. *100*, 15655–15660.

Meurs, E., Chong, K., Galabru, J., Thomas, N.S., Kerr, I.M., Williams, B.R., and Hovanessian, A.G. (1990). Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon. Cell *62*, 379–390.

Meyerson, N.R., and Sawyer, S.L. (2011). Two-stepping through time: mammals and viruses. Trends Microbiol. *19*, 286–294.

Morens, D.M., Folkers, G.K., and Fauci, A.S. (2004). The challenge of emerging and re-emerging infectious diseases. Nature *430*, 242–249.

Moss, B. (2013a). Poxviridae: the viruses and their replication. In Fields Virology, D.M. Knipe and P.M. Howley, eds. (Lippincott Williams & Wilkins), 2129–2159.

Moss, B. (2013b). Poxvirus DNA replication. Cold Spring Harb. Perspect. Biol. *5*,

a010199.

Moss, B. (2012). Poxvirus cell entry: how many proteins does it take? Viruses *4*, 688–707.

Moss, B. (2016). Membrane fusion during poxvirus entry. Semin. Cell Dev. Biol. *60*, 89–96.

Moss, B., and Shisler, J.L. (2001). Immunology 101 at poxvirus U: immune evasion genes. Semin. Immunol. *13*, 59–66.

Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B., Hubisz, M.J., Fledel-Alon, A., Tanenbaum, D.M., Civello, D., White, T.J., J. Sninsky, J., and Adams M.D. (2005). A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. *3*, e170.

Odom, M.R., Hendrickson, R.C., and Lefkowitz, E.J. (2009). Poxvirus protein evolution: family wide assessment of possible horizontal gene transfer events. Virus Res. *144*, 233–249.

Pond, S.L., Frost, S.D., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. Bioinformatics *21*, 676-679.

Potrykus, J., Ballou, E.R., Childers, D.S., and Brown, A.J. (2014). Conflicting interests in the pathogen–host tug of war: fungal micronutrient scavenging versus mammalian nutritional immunity. PLoS Pathog. *10*, e1003910.

Roberts, K.L., and Smith, G.L. (2008). Vaccinia virus morphogenesis and dissemination. Trends Microbiol. *16*, 472–479.

Rothenburg, S., Seo, E.J., Gibbs, J.S., Dever, T.E., and Dittmar, K. (2009). Rapid evolution of protein kinase PKR alters sensitivity to viral inhibitors. Nat. Struct. Mol. Biol. *16*, 63–70.

Sawyer, S.L., Wu, L.I., Emerman, M., and Malik, H.S. (2005). Positive selection of primate TRIM5α identifies a critical species-specific retroviral restriction domain. Proc. Natl. Acad. Sci. U.S.A. *102*, 2832–2837.

Schmidt, F.I., Bleck, C.K., and Mercer, J. (2012). Poxvirus host cell entry. Curr. Opin. Virol. *2*, 20–27.

Seet, B.T., Johnston, J.B., Brunetti, C.R., Barrett, J.W., Everett, H., Cameron, C., Sypula, J., Nazarian, S.H., Lucas, A., and McFadden, G. (2003). Poxviruses and immune evasion. Annu. Rev. Immunol. *21*, 377–423.

Shackelton, L.A., Parrish, C.R., Truyen, U., and Holmes, E.C. (2005). High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc. Natl. Acad. Sci. U.S.A. *102*, 379–384.

Smith, G.L., and Law, M. (2004). The exit of vaccinia virus from infected cells. Virus Res. *106*, 189–197.

Sniegowski, P.D., Gerrish, P.J., Johnson, T., and Shaver, A. (2000). The evolution of mutation rates: separating causes from consequences. Bioessays *22*, 1057–1066.

Stapleford, K.A., Coffey, L.L., Lay, S., Bordería, A.V., Duong, V., Isakov, O., Rozen-Gagnon, K., Arias-Goeta, C., Blanc, H., Beaucourt, S., Haliloglu, T., Schmitt, C., Bonne, I., Ben-Tal, N., Shomron, N., Failloux, A.B., Buchy, P., and Vignuzzi, M. (2014). Emergence and transmission of arbovirus evolutionary intermediates with epidemic potential. Cell Host Microbe *15*, 706–716.

Stremlau, M., Perron, M., Welikala, S., and Sodroski, J. (2005). Species-specific bariation in the B30.2(SPRY) domain of TRIM5alpha determines the potency of human immunodeficiency virus restriction. J. Virol. *79*, 3139–3145.

tenOever, B.R. (2016). The evolution of antiviral defense systems. Cell Host Microbe *19*, 142–149.

Upton, C., Slack, S., Hunter, A.L., Ehlers, A., and Roper, R.L. (2003). Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. J. Virol. *77*, 7590–7600.

Van Valen, L. (1973). A new evolutionary law. Evol. Theor. *1*, 1–30.

Vasilakis, N., Deardorff, E.R., Kenney, J.L., Rossi, S.L., Hanley, K.A., and Weaver, S.C. (2009). Mosquitoes put the brake on arbovirus evolution: experimental evolution reveals slower mutation accumulation in mosquito than vertebrate cells. PLoS Pathog. *5*, e1000467.

Watson, J.C., Chang, H.W., and Jacobs, B.L. (1991). Characterization of a vaccinia virus-encoded double-stranded RNA-binding protein that may be involved in inhibition of the double-stranded RNA-dependent protein kinase. Virology *185*, 206–216.

Weber, F., Wagner, V., Rasmussen, S.B., Hartmann, R., and Paludan, S.R. (2006). Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. J. Virol. *80*, 5059–5064.

Werden, S.J., Rahman, M.M., and McFadden, G. (2008). Poxvirus host range genes. Adv. Virus Res. *71*, 135–171.

Wlasiuk, G., and Nachman, M.W. (2010). Adaptation and constraint at Toll-like receptors in primates. Mol. Biol. Evol. *27*, 2172–2186.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586-1591.

Yang, Z., Maruri-Avidal, L., Sisler, J., Stuart, C.A., and Moss, B. (2013). Cascade

regulation of vaccinia virus gene expression is modulated by multistage promoters. Virology *447*, 213–220.

Yap, M.W., Nisole, S., and Stoye, J.P. (2005). A single amino acid change in the SPRY domain of human Trim5alpha leads to HIV-1 restriction. Curr. Biol. *15*, 73–78.

Zhang, C.Y., Wei, J.F., and He, S.H. (2006). Adaptive evolution of the spike gene of SARS coronavirus: changes in positively selected sites in different epidemic groups. BMC Microbiol. *6*, 88.

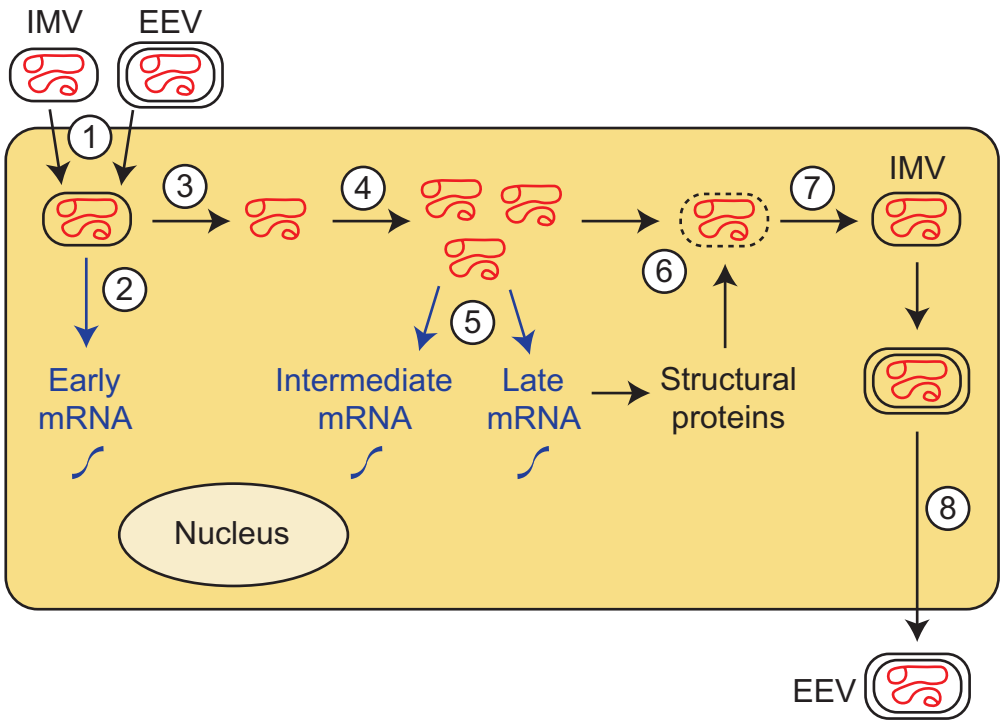


Figure 1.1. Poxvirus replication cycle. Poxvirus replication takes place in the cytoplasm of host cells. Infection initiates with binding and entry of infectious IMV or EEV virions (1). Early transcription occurs within the capsid (2), followed by uncoating of the viral genome (3) and DNA replication (4). Post-replicative transcription of intermediate and late genes follows in a temporal manner (5), producing structural proteins required for virion assembly (6) and maturation (7) to form infectious IMV particles. These virions can be further processed and transported to the cell membrane, where they are released as infectious EEV virions (8).

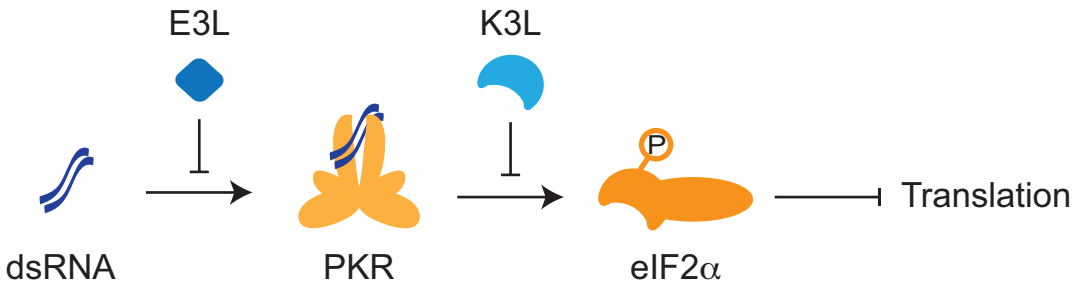


Figure 1.2. Vaccinia virus inhibits the PKR pathway. Double-stranded RNA produced during the viral lifecycle is detected by the host nucleic acid sensor PKR. Upon binding dsRNA, PKR becomes activated and phosphorylates its substrate, eIF2α. Phosphorylated eIF2α leads to a global block in protein translation. VACV encodes two inhibitors of the PKR pathway, E3L and K3L. E3L binds dsRNA and prevents PKR activation, while K3L is a mimic of eIF2α.

CHAPTER 2

POXVIRUSES DEPLOY GENOMIC ACCORDIONS TO ADAPT RAPIDLY AGAINST HOST ANTIVIRAL DEFENSES

Reprinted with permission from Elde, N.C., Child, S.J., Eickbush, M.T., Kitzman, J.O., Rogers, K.S., Shendure, J, Geballe, A.P., and Malik, H.S. (2012). Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell *150*. 831-41.

Cell

Poxviruses Deploy Genomic Accordions to Adapt Rapidly against Host Antiviral Defenses

Nels C. Elde,[1,4,*] Stephanie J. Child,[2] Michael T. Eickbush,[1] Jacob O. Kitzman,[5] Kelsey S. Rogers,[4] Jay Shendure,[5] Adam P. Geballe,[2,3,6,7] and Harmit S. Malik[1,8,*]
[1]Division of Basic Sciences
[2]Division of Human Biology
[3]Division of Clinical Research
Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[4]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112, USA
[5]Department of Genome Sciences
[6]Department of Medicine
[7]Department of Microbiology
University of Washington, Seattle, WA 98195, USA
[8]Howard Hughes Medical Institute, Seattle, WA 98109, USA
*Correspondence: nelde@genetics.utah.edu (N.C.E.), hsmalik@fhcrc.org (H.S.M.)
http://dx.doi.org/10.1016/j.cell.2012.05.049

SUMMARY

In contrast to RNA viruses, double-stranded DNA viruses have low mutation rates yet must still adapt rapidly in response to changing host defenses. To determine mechanisms of adaptation, we subjected the model poxvirus vaccinia to serial propagation in human cells, where its antihost factor K3L is maladapted against the antiviral protein kinase R (PKR). Viruses rapidly acquired higher fitness via recurrent K3L gene amplifications, incurring up to 7%–10% increases in genome size. These transient gene expansions were necessary and sufficient to counteract human PKR and facilitated the gain of an adaptive amino acid substitution in K3L that also defeats PKR. Subsequent reductions in gene amplifications offset the costs associated with larger genome size while retaining adaptive substitutions. Our discovery of viral "gene-accordions" explains how poxviruses can rapidly adapt to defeat different host defenses despite low mutation rates and reveals how classical Red Queen conflicts can progress through unrecognized intermediates.

INTRODUCTION

Rapidly diversifying host immune repertoires pose significant barriers to successful transmission of viruses within and between species. Many RNA viruses counter host immunity with high mutation rates coupled with short generation times and large effective population sizes, allowing rapid exploration of mutational space for adaptation. Previous studies have elucidated how the high mutation rates of RNA viruses are required for tropism in animal species (Pfeiffer and Kirkegaard, 2005; Vignuzzi et al., 2006), despite driving them perilously close to an "error catastrophe" threshold (Crotty et al., 2001; Vignuzzi et al., 2005). Much less is known about the adaptive strategies of large double-stranded DNA viruses that, despite short generation times, seem to evolve more slowly than RNA viruses (Drake et al., 1998; Drake and Holland, 1999; Li et al., 2007; Lynch, 2010) yet comprise highly successful lineages.

The poxviruses are large double-stranded DNA viruses that are notable because they can infect most vertebrates (chordopoxvirinae) and insects (entomopoxvirinae) (Harrison et al., 2004; Moss, 2007). Poxviruses encode a large array of genes that hijack or antagonize components of the host cell to promote viral fitness (Bahar et al., 2011; Werden et al., 2008). Strong natural selection at these host-virus interfaces can drive rapid adaptations, which are reflected by the fixation of nonsynonymous substitutions in coding regions. These substitutions can either enhance or weaken interactions between host and viral factors (Emerman and Malik, 2010). The Red Queen hypothesis postulates that such interactions evolve as an ongoing series of counteradaptations (Dawkins and Krebs, 1979; Meyerson and Sawyer, 2011; Van Valen, 1973). Several retrospective evolutionary studies comparing the sequences of open reading frames (ORFs) among poxvirus genomes provide ample evidence of adaptive changes to overcome rapidly evolving host defenses (Bratke and McLysaght, 2008; McLysaght et al., 2003).

One of the most potent innate defenses against viruses is encoded by protein kinase R (PKR), which is activated upon sensing double-stranded RNA that accumulates in the host cytoplasm during many viral infections (Weber et al., 2006). Active PKR phosphorylates the translation initiation factor eIF2α to inhibit protein production, which strongly impairs viral replication (Kaufman, 2000). Antagonism by a variety of viruses over the

Cell

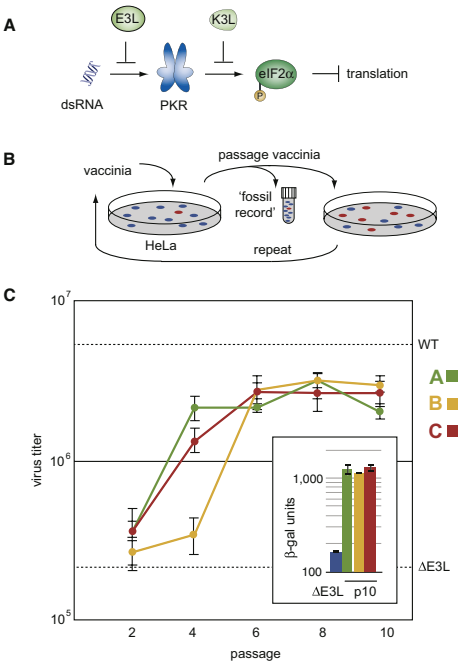


Figure 1. Experimental Evolution of Vaccinia Virus

(A) Vaccinia encodes two host-range factors, E3L and K3L, that inhibit antiviral protein kinase R (PKR).

(B) Vaccinia ΔE3L was repeatedly passaged through HeLa cells in triplicate (see Experimental Procedures). After each infection, virus titer was measured, an aliquot was reserved for the "fossil record," and fresh cells were infected. Red dots represent viruses in which a theoretical adaptation arises in the population and increases in frequency.

(C) With passaging of ΔE3L virus, replicate populations A–C gained the ability to replicate in HeLa cells as judged by viral titers assayed over the course of the experiment. Dotted horizontal lines show the average titer of parental virus (ΔE3L) and the wild-type Copenhagen strain of vaccinia (WT). Because E3L was replaced by lacZ, we could also use β-gal activity assays as a proxy for measuring virus replication (inset). Gains in replication were corroborated by measurements of β-gal activity, shown in arbitrary units, from parental virus and replicates A–C at passage 10 (p10; inset). Data are represented as mean ± SEM.

course of primate evolution has driven the rapid evolution of PKR in primates (Elde et al., 2009; Rothenburg et al., 2009). As a result of this diversification, primate variants of PKR are differentially susceptible to inhibitors encoded by different viruses. Vaccinia virus encodes two such antagonists, K3L and E3L, which each specifically inhibit PKR by distinct mechanisms in different species (Davies et al., 1993; Langland and Jacobs, 2002). The rapid evolution of PKR has led to species-specific resistance to K3L among primates.

A partial explanation for flexibility in poxvirus host range is provided by several "host-range" factors, like K3L and E3L, which can overcome species-specific blocks to infection. Left unanswered, however, is the fundamental question of how poxviruses, despite their low mutation rate, efficiently explore mutational space and overcome rapidly evolving immune factors, such as PKR, as they move between host species. For example, some poxviruses, such as cowpox, vaccinia, and monkeypox, infect both animals and humans in zoonotic infections, with the latter cases being of considerable biomedical importance. The fact that these viruses can infect divergent host species despite diverse mechanisms of immunity suggests that they employ different means of evolutionary adaptation from those known for rapidly evolving RNA viruses to overcome low mutation rates. Moreover, even within the same host populations, poxviruses must adapt rapidly to persist despite low mutation rates. Gaining insight into the currently unknown basis of adaptation in this biomedically and evolutionarily important class of viruses is essential for understanding how these viruses infect a wide range of hosts.

Fundamental insights into evolutionary mechanisms have been derived from laboratory and experimental evolution of various microbial populations (Beaumont et al., 2009; Meyer et al., 2012; Montville et al., 2005; Vignuzzi et al., 2006; Wichman et al., 1999). To date, however, nearly all protocols involving experimental evolution of viruses have used strains with high mutation rates (e.g., small RNA viruses) and/or ones that quickly generate large effective population sizes under laboratory conditions (e.g., bacteriophages). Here, we devised an experimental evolution protocol to determine how poxviruses adapt using serial vaccinia infections at relatively small population sizes. By passaging a vaccinia strain deleted for the host range gene E3L, we mimicked a host-switching event, which placed strong selective pressure on K3L to inhibit the antiviral PKR. Our studies reveal mechanisms of poxvirus adaptation to defeat host defenses via rapid episodes of transient gene amplification. This simple and recurrent mechanism of adaptation on a genomic scale also reveals a new general model for mitigating fitness costs associated with large changes in genome structure. Our experiments not only reveal an important means by which poxviruses adapt but also highlight a mechanism by which radical adaptations might be explored in many organisms that have large genomes and low mutation rates.

RESULTS

Experimental Evolution of Vaccinia Virus Reveals Adaptation by Gene Expansion

We used the fact that K3L from vaccinia virus is a poor inhibitor of human PKR to determine how poxviruses can adapt to overcome an initially insurmountable host defense. By starting with the Copenhagen strain of vaccinia virus deleted for E3L (ΔE3L) (Beattie et al., 1995), we placed strong selective pressure on K3L (Figure 1A). Three replicate virus populations were repeatedly propagated in HeLa cells at a low multiplicity of infection of 0.1 plaque-forming units/cell. After 48 hr of infection, we collected the virus, measured titers, and repeated the protocol for ten passages (Figure 1B and Experimental Procedures).

ΔE3L virus initially replicated much less efficiently than wild-type virus in HeLa cells, consistent with previous reports (Langland and Jacobs, 2002). By six rounds of passaging in HeLa cells, each of three replicate virus populations consistently produced 10-fold more infectious viruses than the parental virus strain as measured by plaque assays (Figure 1C) and β-gal activity assays (Figure 1C, inset). Further gains in virus production were not evident during the remaining four passages. Thus, despite a low mutation rate, and relatively small effective population sizes starting at 5×10^5 viruses/passage, all three replicates of vaccinia rapidly gained increased fitness over the course of passaging.

To determine the genetic basis of rapid adaptation, we performed deep sequencing of all virus populations to approximately 1,000-fold genome coverage. We found only two nonsynonymous and one synonymous polymorphism in the genomes of passage 10 virus replicates at a frequency higher than 1% (Figure 2A). This paucity of mutations highlights the low rate of nucleotide substitution in vaccinia genomes. In addition, none of the substitutions were found at high frequency or in all replicates. In contrast, by scanning sequencing coverage among mapped reads, we found a substantial spike in relative depth that mapped exclusively over the K3L gene, which is consistent with a striking increase in copy number of the K3L locus in all three replicate strains (Figure 2A). The increased read-depth corresponds to an average of three (replicates A, C) to four (replicate B) copies of K3L per viral genome (Figure 2A).

To further determine the character of these K3L expansions, we performed Southern blot analysis of viral genomes sampled from points throughout the passaging protocol, using the K3L gene as a probe. These analyses revealed a highly heterogeneous pool of viral genomes in each replicate, with some harboring 15 or more copies of K3L by passage 10 (Figure 2B). Intriguingly, the appearance of multiple copies of K3L paralleled the pronounced increases in virus replication, which occurred by passage 4 for both replicates A and C, and passage 6 for replicate B (compare Figures 1C and 2B). Therefore, increased copy number of K3L directly correlated with increased viral fitness. Additional analysis of discordantly mapping paired-end reads revealed a complex pattern of duplications at the K3L locus exclusive to the adapted strains (Figure 2C; Figure S1 available online; Table S1). Notably, analysis of the duplication break points confirmed the independent origins of the K3L duplications in each of the three replicate populations of adapted vaccinia viruses.

Resequencing of parental genomes to nearly 1,000-fold coverage failed to reveal any pre-existing duplications of K3L in this homogenous virus population (Table S1). PCR confirmed the independent emergence of tandem duplications of K3L at several distinct break points in the genes adjacent to K3L in the evolved populations. In addition to our initial experiment, three additional ΔE3L virus replicates passaged under the same conditions all yielded expansions of K3L as judged by PCR to detect K3L-associated break points (Figure S2), demonstrating the reproducibility of gene amplification as a means of adaptation.

To determine if genomic amplification of K3L accounted for enhanced viral replication, we first compared K3L protein expression between parental and passaged viruses (Figure 3A). Each passage 10 replicate expressed considerably more K3L protein than even wild-type virus, which still replicates more efficiently than any of the adapted strains because of the presence of E3L (Figure 3A). To determine whether overexpression of K3L alone is sufficient to explain the robust increase in viral fitness, we specifically knocked down K3L expression using siRNA transfections. Successful knockdown of K3L reduced virus replication to levels indistinguishable from prepassaged, parental ΔE3L virus (Figure 3B). Therefore, K3L copy number amplification was necessary and sufficient for increased expression of K3L, which caused the substantial gain in virus fitness we observed. Because K3L mimics PKR's substrate eIF2α and inhibits PKR by direct binding (Dar and Sicheri, 2002), it appears that an increased abundance of even weakly active K3L antagonizes PKR by mass action, providing the basis for an immediate evolutionary advantage due to genomic expansion.

Acquisition of an Adaptive Substitution following Gene Expansion

Whereas K3L copy number amplification provided a rapid, recurrent basis for adaptation in all of the replicate populations, most studies of virus evolution identify adaptive point mutations as the basis of fitness gains. We found that two of the replicates from our study possessed a single amino acid substitution, H47R, in K3L at frequencies of 3% and 12% (Table S2). Remarkably, this exact substitution was previously identified in a genetic screen for K3L mutants with enhanced activity against human PKR using a yeast assay interrogating the K3L-PKR interaction (Kawagishi-Kobayashi et al., 1997). We isolated this variant from the adapted passage 10 replicate C vaccinia population containing a single copy of K3L with the H47R mutation and tested its fitness relative to viruses with gene expansions of wild-type K3L sequence. Both the single copy H47R-bearing virus and the multicopy K3L virus had marked improvements in fitness over the parental single-copy wild-type K3L virus (Figure 3C). These findings also strongly support the idea that fitness gains realized by K3L amplification and acquisition of H47R in serially propagated virus populations reflect specific antagonism of PKR. In addition, although only 12% of the K3L genes at passage 10 of replicate C contain the H47R mutation (541 reads out of 4,146 total sequencing reads), 47% of single-copy K3L genes from the same virus population harbor this substitution (24 cases with H47R out of 51 K3Ls sampled via PCR and sequencing; $p < 0.0001$, Fisher's exact test). This result implies that low copy number viruses are strongly selected to possess the adaptive H47R allele of K3L compared to K3L expanded viruses in the same pool. Thus, our experimental evolution strategy uncovered two responses to the same selective challenge: adaptive copy number expansions seen in all replicate populations and an adaptive nonsynonymous mutation seen at low frequency in two of three replicates.

The emergence of the H47R substitution raised the possibility that K3L expansions facilitated the appearance of the variant by providing additional targets for mutation to generate increased variation, including beneficial substitutions, like H47R. Examination of parental ΔE3L strain sequences mapping to codon 47 of the K3L coding region failed to reveal H47R in the starting

Cell

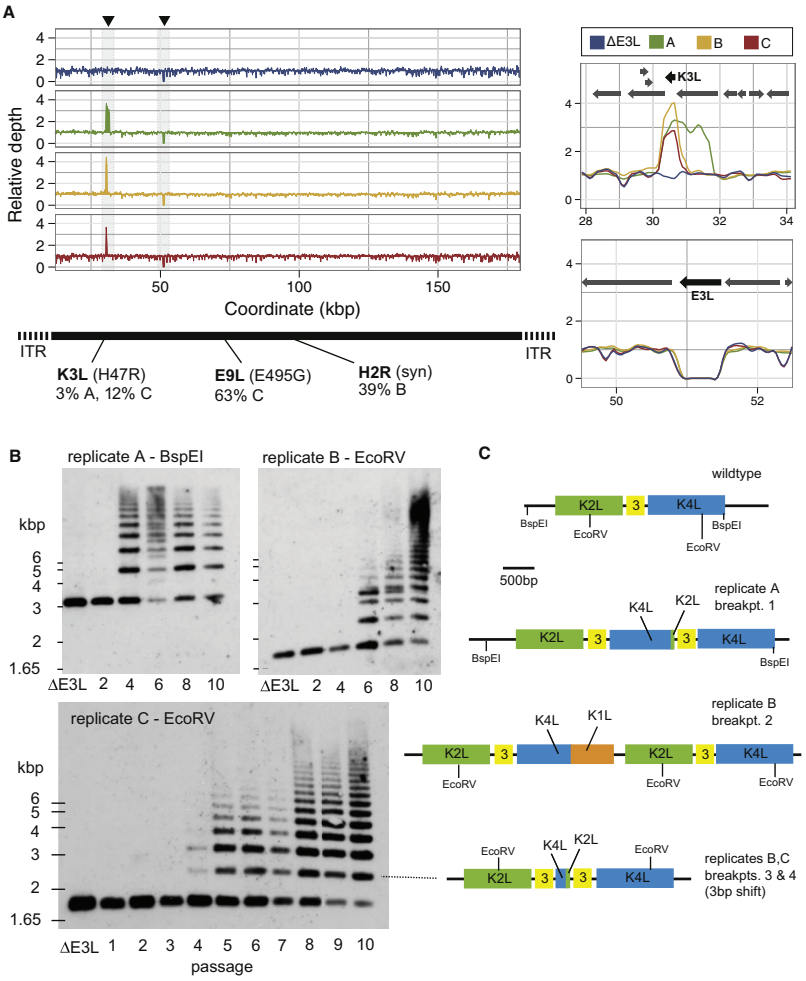


Figure 2. Rapid Evolution of Copy Number Amplification

(A) Genomic DNA traces of parent and virus replicates A–C at passage 10, showing relative depth of sequence coverage across the entire ~180 kb vaccinia genome, except for the inverted terminal repeat regions (ITR; ITRs are shown in Figure S4). Increased K3L sequence coverage was observed in replicates A–C (upper-right; also see Figures S1 and S2), whereas coverage of E3L was missing in all populations (lower-right). Polymorphisms were detected in the K3L, E9L, and H2R genes as indicated (also see Table S2). We also detected 15 nucleotide differences between the parent and reference Copenhagen strain, which were unrelated to our protocol (Goebel et al., 1990) (Table S6).

(B) Southern blots of the K3L locus of viral replicates A, B, and C during the course of experimental evolution from passages 1–10 reveal rapid gene copy number amplification of K3L. Amplification may be a consequence of recombination-driven tandem duplications (Figure S3).

(C) Diagrammed is the parental K3L locus and the major duplication break points near K3L that led to copy number amplification (also see Table S1). We also found duplications at the highly repetitive termini of the vaccinia genomes, consistent with previous observations (Moss et al., 1981) (see Figure S4). However, these terminal expansions were also present in the parental strain and therefore not a consequence of our experiments.

Cell

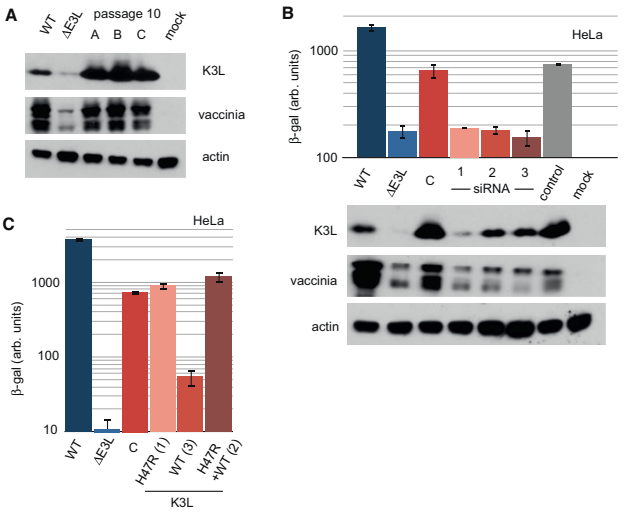


Figure 3. Overexpression of K3L Protein in Adapted Strains Is Necessary and Sufficient for Increased Viral Fitness

(A) Western blot analysis of vaccinia and mock-infected HeLa cells with wild-type (WT), parental (ΔE3L), and virus replicates A–C at passage 10 for levels of K3L, vaccinia proteins, and actin.

(B) siRNA knock down of K3L in infected HeLa cells. Viral replication was measured by β-gal activity assay of virus and mock-infected cells transfected with K3L-specific siRNAs (1–3) or a nontargeting control. Western blot shows levels of K3L, vaccinia proteins, and actin.

(C) β-gal activity from HeLa cells infected with various vaccinia viruses. Wild-type vaccinia virus, ΔE3L parental virus, and passage 10 virus from replicate C were compared to plaque-purified clones with either a single copy H47R-K3L (1), a single copy WT-K3L (3), or a multicopy K3L virus with both WT and the H47R substitutions (2). Numbers correspond to plaque-purified viruses described in Figure 6. Data are represented as mean ± SEM.

population (Table S2). Also consistent with K3L expansions preceding the acquisition of H47R was our observation that replicate B, the slowest population to undergo K3L amplification (Figure 1C), lacked any detectable H47R K3L alleles by passage 10. Together, these data suggest that H47R appeared after K3L gene expansion.

To determine when H47R emerged during the course of passaging, we amplified K3L from genomic samples of replicate C at each passage point and sequenced amplicon populations. K3L expansions, as judged by PCR amplification across K3L duplications, preceded the appearance of the H47R substitution during viral passaging (Figure 4). Sequencing of PCR products spanning K3L duplications at late passage points revealed the presence of H47R in both single and multicopy K3L arrays. We further sampled the composition of multicopy arrays by sequencing 34 clones bearing two K3L gene copies. Of these, ten had both wild-type and H47R-containing copies of K3L. Together, these data strongly suggest that K3L expansions facilitated the emergence and increased frequency of the H47R substitution. Thus, not only are the K3L gene expansions immediately advantageous because of their ability to provide enhanced antagonism of human PKR but they also facilitate the acquisition of single Darwinian ''gain-of-function'' changes by increasing the number of mutational targets despite an otherwise modest mutation rate.

A General Mechanism of Gene Expansion in Vaccinia Genomes

Two previously proposed mechanisms could account for facile gene expansion in vaccinia genomes. One study proposed unequal crossover as a mechanism to account for variation in the distal inverted terminal repeat regions of the vaccinia genome (Moss et al., 1981). A second study pointed to the possibility that gene expansions could occur episomally, because vaccinia can promote rapid amplification of nonviral plasmids. To determine whether the K3L locus was being expanded by a mechanism involving episomes containing K3L, we used RecBCD enzyme, a complex helicase and nuclease, which rapidly digests all linear genomic DNA but spares any potential circular episomes (Eichler and Lehman, 1977; Smith et al., 1995). Southern blot analysis of undigested genomes, as well as those treated with RecBCD to eliminate linear DNA forms that might have otherwise obscured circular forms, failed to reveal the presence of circular DNA intermediates containing K3L in genomes of viruses passaged in HeLa cells (Figure S3). Thus, we can exclude an episomal mode of amplification, which suggests that recombination-driven K3L gene expansions within linear genomes of each replicate virus account for the rapid adaptation of vaccinia virus.

Recombination-mediated gene expansions, which can facilitate the acquisition of adaptive substitutions, could represent a general mechanism for rapid adaptation of poxviruses. To determine if gene duplications similar to K3L were present at low frequency in virus populations, we assayed for the presence of additional duplications in virus populations. Twelve sets of PCR primers were designed to amplify products across potential duplication break points at locations spaced roughly evenly throughout the genome (Table S3). These probes for duplications were used to query three vaccinia strains (Western Reserve, Copenhagen-VC2, and the parental strain used in this study, Copenhagen-VC2-ΔE3L). Despite our limited sampling, we found that each strain harbored at least one duplication break point in virus populations (Figure 5). We did not detect any shared sequence motifs or consistent patterns of homology at sites in which duplications arose that might promote genomic expansions. However, each of the duplications we uncovered

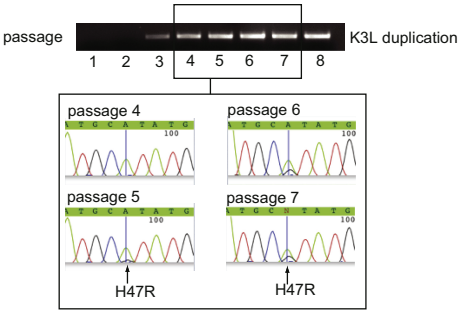


Figure 4. K3L Expansions Precede the Appearance of the H47R Substitution

PCR with genomic samples from the "fossil record" of replicate C with primers amplifying duplications of K3L is consistent with Southern blot analysis. Sequencing of duplicated K3L regions reveals that the appearance of the mutation leading to the H47R substitution is only appreciable after duplications of K3L appeared.

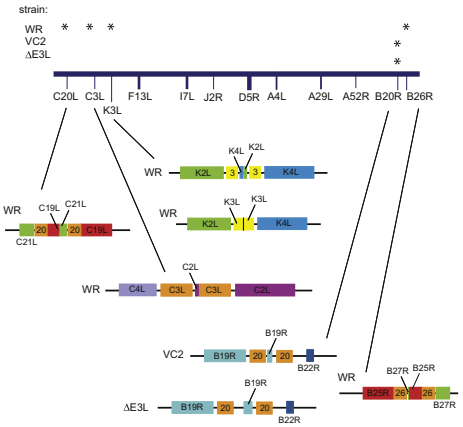


Figure 5. Low-Frequency Gene Duplications in Vaccinia Genomes

Of the primers tested (Table S3), duplications at several loci near the termini of the genome were amplified and sequenced (Table S4). Break-point regions are drawn to scale, highlighting the location of each duplication point. One break point observed in K3L leads to a fusion product in place of the wild-type copy. The C20L and B26R regions are mirrored at each end of the chromosome, such that the reported break points could be at either end of the genome. Additional low-frequency duplications were detected in deep sequencing reads of parental and adapted virus genomes (Table S5).

maps toward the distal regions of the vaccinia genome, which have a higher frequency of recombination and are comprised of many genes that are dedicated to the modulation of host factors (Moss, 2007). In contrast, genes toward the center of the vaccinia genome are essential for virus-specific functions in the cytoplasm of the host cell. Four of the twelve primer sets tested amplified duplications in the Western Reserve strain, which included several break points from duplications of K3L (Table S4). Further examination of deep sequencing reads also revealed rare duplication break points that were confirmed for two cases by PCR in populations of adapted strains at genomic locations outside the K3L locus (Table S5). These data reveal that virus populations contain a variety of low-frequency break points that could rapidly seed genomic expansions under suitable selective conditions, such as changing hosts. Much like random nucleotide substitutions that can start adaptive sweeps driven by selection, we propose that these low-frequency duplication events across the vaccinia genome can drive rapid gene expansions to overcome host defenses, followed by expanded sampling of point mutations. Given the low rate of nucleotide substitutions uncovered by our sequencing analyses, the frequency of these duplications may be on par with or exceed the frequency of nucleotide substitutions, as has been previously observed in some eukaryotic genomes (Lipinski et al., 2011; Lynch et al., 2008).

Genomic Accordions Reveal Tradeoffs Associated with Genomic Expansion

K3L copy number expansions represent a significant increase in vaccinia genome size. Although comprising a minor fraction of the population, viruses harboring 15-fold amplifications of K3L have genomes 7%–10% larger than the parental strain (Figure 2B). Vaccinia can package similarly sized genomic additions with a seemingly negligible impact on replication efficiency (Smith and Moss, 1983), but poxvirus genomes also undergo adaptive size reductions under a variety of conditions (Hendrickson et al., 2010; Meisinger-Henschel et al., 2007). Therefore, to determine if there was an evolutionary tradeoff associated with large genomic expansions of K3L, we removed selective pressure from amplifications of K3L. Unlike in human cells, single copy K3L from vaccinia virus defeats PKR in BHK (hamster) cells (Figure 6A) (Beattie et al., 1995; Langland and Jacobs, 2002); therefore, passaging virus in these permissive cells directly tests whether the expansions incurred a fitness tradeoff.

We plaque-purified nine vaccinia clones from the population of passage 10 viruses from replicate C by passaging four times in BHK cells. Southern analyses of these viral genomes revealed that these viruses contained either only a single copy of K3L or nearly identical multicopy arrays with considerably fewer copies of K3L compared to passage 10 viruses from HeLa cells (Figure 6B). To rule out the possibility that secondary genetic changes had caused contraction of the K3L expansions, we reintroduced selective pressure on one of the plaque-purified multicopy K3L viruses (5*) with four additional passages in HeLa cells. Reapplication of selective pressure led to a clear increase in K3L copy number compared with passaging the same viruses in BHK cells (Figure 6C). Thus, removing the selective pressure that led to gene amplification resulted in a reproducible and reversible reduction in K3L locus expansion, consistent with there being a fitness cost associated with the genomic expansions. Alternatively, it is possible that simply relaxing

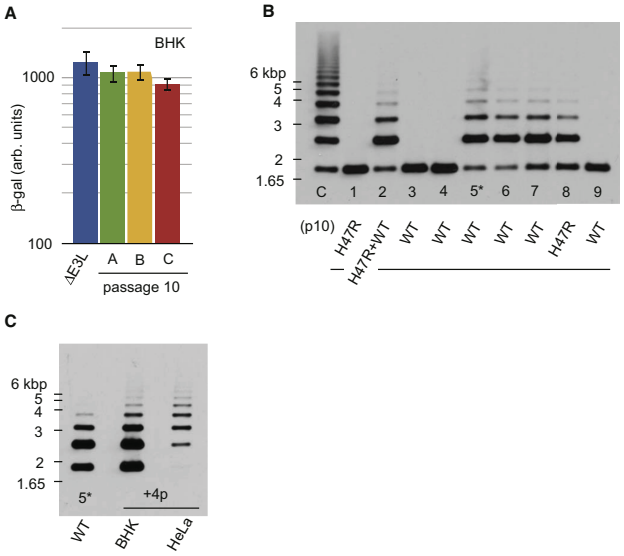


Figure 6. Evolution of Transient Copy Number Amplification

(A) β-gal activity from BHK cells infected with parent virus (ΔE3L) and replicates A–C at passage 10. Data are represented as mean ± SEM.

(B) Southern blot of the K3L locus from plaque-purified vaccinia viruses obtained from serial plaque purification of passage 10 replicate C virus in BHK cells. The presence or absence of the H47R substitution in K3L is indicated for each strain.

(C) Southern blot of multicopy plaque-purified virus (clone 5* from B) after four additional passages (+4p) in BHK or HeLa cells. Viruses passaged in HeLa cells showed re-expanded K3L copy number relative to those passaged in BHK cells.

selective pressure for K3L expansion leads to a relative loss of K3L because of a bias in which recombination-driven losses exceed gains. It is notable, however, that replicate B virus populations lacking the H47R substitution continue to show ongoing increases in K3L copy number at passage 10 as judged by Southern analysis (Figure 2B). This suggests that copy number of K3L may continue to expand until a beneficial mutation provides an adaptive alternative to genomic expansion.

DISCUSSION

Our experimental evolution studies of vaccinia evoke a "gene accordion" model for genome evolution. Localized genomic regions transiently expand to enhance levels of gene expression, while simultaneously providing additional gene copies for sampling mutational space (Figure 7). Subsequent contractions mitigate the costs of replicating larger genomes while retaining adaptive alleles like H47R, which arose during the expansion. Thus, a simple mechanism of recombination-driven genomic expansions and contractions facilitates the rapid evolution of virus populations with otherwise low mutation rates.

Gene accordions do not appear unique to the K3L locus. In addition to the low-frequency duplications we detected in vaccinia genomes (Figure 5; Table S4), this model of adaptation through intermediates of transient gene expansion also helps explain a variety of previous observations in the genomes of large double-stranded DNA viruses, including ancient gene family expansions among several poxviruses (McLysaght et al., 2003) and herpesviruses (Searles et al., 1999), chemically induced gene amplification in vaccinia (Slabaugh and Mathews, 1986; Slabaugh et al., 1989), and a potentially adaptive duplication observed in the myxoma poxvirus (Kerr et al., 2010). A particularly clear example of such adaptive gene expansions is evident in the genomes of avipoxviruses, which devote as much as 30% of the genome to a small set of large gene families that includes many ankyrin repeat-containing genes (Afonso et al., 2000). Avipoxviruses illustrate an exceptional example of gene duplications facilitating neofunctionalization of individual copies of the amplified gene family, leading to the retention of multiple variants rather than contraction back to a single copy. We propose that sampling a reservoir of low-frequency duplications at various genomic regions in the virus population reflects an underappreciated but common mechanism seeding dynamic and adaptive gene expansions in these viruses. Our observations reveal that such viruses adapt to new, hostile host environments first by rapid gene expansion rather than gene mutation.

These results also reveal that a strength of the experimental evolution strategy is its power to reveal important evolutionary intermediates that could be too fleeting to capture by less frequent sampling or missed all together by studying viruses long after host switches. For example, we could clearly demonstrate that vaccinia genomes underwent a specific K3L gene expansion that exactly correlated with all fitness gains observed during our experiment. However, by ten passages the K3L gene had already been replaced by the H47R variant in 12% of all K3L genes in one of the replicates. Sampling at a later point could conceivably obscure the accordion-like origins of an H47R-like adaptation as the variant sweeps through the population and replaces large and costlier gene expansions. This duality of adaptation might also be reflected in the different stages of viral transmission between species, including zoonosis. Gene expansions might be favored early in infection when most viral adaptation is centered on overcoming host defenses, as was modeled by our experiment. At later stages of infection, gene expansions are likely to be strongly selected against as viruses compete not only against the initial challenge imposed by host defenses but also against larger populations of fit viruses carrying the

Cell

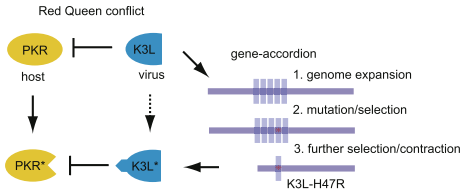


Figure 7. Genomic Accordion Model of Poxvirus Evolution
The model depicts Red Queen conflicts proceeding through "gene-accordion" intermediates. Gene copy number amplification provides an immediate fitness advantage and additional gene copies for sampling potentially beneficial mutations. After acquisition of a beneficial mutation, virus genomes are selected that retain advantageous point mutations without the gene expansion, leading to contractions down to a single gene copy.

equivalent of an H47R mutation. Future studies will be important to assess not only the generality of this adaptive mechanism but also to compare the dynamics observed in our experiments to those that occur during natural poxviral infections.

Despite the rapid gene expansions observed in our study, very few poxviruses sequenced from the wild have been found to harbor recent gene duplications or expansions, with the ankyrin repeat genes in avipoxviruses being a clear exception. Although some expansions may be obscured by subsequent contractions, it is also possible that many expansion events have been missed because of an ascertainment bias. Until recently, sequencing methodologies rarely accounted for copy number variation as a source of genetic heterogeneity. Indeed, sequencing of different strains of variola (smallpox) virus with the aim of determining how variola major had a higher fatality rate than variola minor focused primarily on nucleotide substitutions, which revealed less than 2% divergence (Esposito et al., 2006). Although this genetic divergence might be sufficient to explain differential pathogenicity, important instances of copy number variation might have gone undetected in these comparisons. New sequencing technologies greatly increase the prospects of capturing such intermediate states from poxviruses that have undergone minimal passaging in the lab.

Our study also highlights the important role that gene copy number expansion plays in adaptation of genomes with low mutation rates. This mode of adaptation may be especially prevalent when an immediate advantage can be gained via mass action. Examples include selection for resistance to drug therapies against cancer (Drummond et al., 1997; Schimke, 1986), for enhanced enzymatic degradation of toxins (Sandegren and Andersson, 2009), or for overcoming a host antiviral protein, like PKR. In these cases, the duplication event is itself nonneutral and can be further acted upon by recombination to drastically alter genome architecture. Whereas rapid evolution of copy number variation is known for some nonvirus microbial populations, including *Plasmodium*, yeast, and a variety of bacteria (Bergthorsson et al., 2007; Demuth and Hahn, 2009; Dunham et al., 2002; Kugelberg et al., 2010; Nair et al., 2008; Pränting and Andersson, 2011; Stambuk et al., 2009; Sun et al., 2009), adaptation via gene expansion in poxviruses represents an unappreciated mechanism of evolution in this large, medically relevant class of DNA viruses (Harrison et al., 2004). In the cases cited above, copy number expansions may also be poised for adaptive contractions that unfold on longer time scales, thus obscuring the "genomic-accordion" mechanism revealed here by vaccinia virus (but also see Pränting and Andersson, 2011).

Some of the passage 10 viruses we assayed have a 7%–10% expansion in genome size purely by virtue of a single gene expansion. In a 200 kb genome, this is a noteworthy adaptation for its major alteration of genomic structure. Our findings also indicate that a diversity of adaptive strategies—copy number variation in addition to nucleotide substitution—falls well within the viral kingdom. It seems likely that this mode of adaptation is applicable to many other viruses that have similarly low mutation rates and relatively limited restrictions on genome size. Exceptions are likely to be viruses that have low mutation rates but strict limits on genome size because of a highly constrained capsid structure (e.g., adenoviruses and certain bacteriophages). It will be informative to determine whether other large DNA viruses, like herpesviruses, which replicate in the nucleus and also show evidence of gene expansions (Searles et al., 1999), follow similar gene-accordion dynamics as the poxviruses, which replicate in the cytoplasm of host cells.

There are striking parallels between the challenges imposed on viruses by exposure to new batteries of innate defense genes and the challenges imposed on organisms adapting to strong selective pressure, like bacteria grown under stress-inducing conditions (Andersson et al., 1998; Cairns et al., 1988; Foster, 1994; Hastings et al., 2000; Hendrickson et al., 2002). In each of these cases, a significant step in the adaptive process involves rapid and short-lived gene expansions that permit the selection of "escape" variants. Our study thus reveals a previously unappreciated insight into the dynamics of Red Queen conflicts between pathogens and hosts (Elde and Malik, 2009; Emerman and Malik, 2010; Meyerson and Sawyer, 2011). Traditionally viewed primarily as substrates for rapid selection of nonsynonymous codon substitutions, these interactions may also depend on adaptive gene amplifications that are overlooked because of their transience or difficulty of detection (Figure 7). Indeed, "gene-accordions" may be a critical feature of many genetic conflicts involving strong selection.

EXPERIMENTAL PROCEDURES

Viruses and Reagents

The VC2 (wild-type) Copenhagen strain of vaccinia virus and the E3L deletion mutant (ΔE3L), with the lacZ gene replacing E3L, were a gift from B. Jacobs (Beattie et al., 1995). HeLa and BHK cells were cultivated in Dulbecco's modified Eagle's medium (DMEM) + 10% Nu serum (BD Biosciences, Franklin Lakes, NJ, USA).

Experimental Evolution of Vaccinia Virus

For each passage of vaccinia, 100 mm dishes were seeded with a fresh aliquot of HeLa cells in triplicate (5×10^6 cells/dish). Cells were infected with ΔE3L virus at a multiplicity of infection of 0.1 pfu/cell. After 48 hr, infected cells were pelleted and washed, and viruses were harvested by three freeze/thaw cycles in a mixture of dry ice/ethanol and were resuspended in 1 ml of

Cell

DMEM + 10% Nu serum. Viral titers were calculated by performing plaque assays in BHK cells between each passage. After ten passages, aliquots of virus reserved from each passage were expanded with a single 48 hr infection of BHK cells and viral titers were determined. To evaluate the replication properties of these viruses, triplicate wells of HeLa cells were infected at 0.1 pfu/cell using viruses from every other passage and viral yield was measured by titering cell lysates made at 48 hr postinfection. β-gal activity was determined with 4-methylumbelliferyl-β-D-galactopyranoside (MUG) fluorescence assays (Hakki and Geballe, 2005). Genomic DNA was prepared as previously described (Esposito et al., 1981) from viruses obtained at every passage of the protocol.

Deep Genome Sequencing of Vaccinia Strains

Illumina sequencing libraries were prepared from genomic DNA from parental ΔE3L virus and each replicate (A–C) at passage 10 using the Nextera sequencing library construction kit (Epicentre Biotechnologies [Illumina, Inc.], Madison, WI, USA). Multiplexing barcodes were added by PCR as previously described (Adey et al., 2010). Barcoded libraries were combined and sequenced across three lanes on Illumina GA-IIx and HiSeq 2000 instruments. Reads were mapped to the vaccinia reference genome with bwa version 0.6.1 (Li and Durbin, 2009); custom python scripts were used to calculate read depth and find break-point junctions.

Plaque Purification of Vaccinia

BHK cells were infected with dilutions of passage 10, replicate C virus. At ~2 days postinfection, cells were collected from isolated individual plaques, diluted into medium, and lysed by three cycles of freezing and thawing. Dilutions of these first round plaque-purified viruses were used to infect fresh BHK cells, and the process was repeated three more times. After the fourth round of plaque-picking, a stock of each plaque-purified virus was prepared in BHK cells and was used for infections to assess replication and to make viral DNA.

RecBCD Digests

Vaccinia replicate C genomes harvested after 10 passages were first digested with restriction enzymes recognizing sites near K3L break points (EcoRV), within K3L (BstZ17I), at sites distant from K3L (SacI), or left intact. Samples were incubated with five units of RecBCD ±125 μM ATP based on a previously reported method (Eichler and Lehman, 1977) but with a modified reaction buffer containing 1 mM DTT, 20 mM potassium phosphate, and 10% glycerol. Samples were incubated twice for 2 hr at 37°C with five additional units of RecBCD added between incubations. Genomes were resolved by agarose gel electrophoresis at 80 V for 10 hr and Southern blots were performed as described in the methods.

Southern and Western Blot Analysis

For Southern blots, 2–5 μg of vaccinia genomic DNA was digested with EcoRV or BspEI (New England Biolabs, Ipswich, MA, USA). Genomes were resolved by agarose gel electrophoresis, DNA was transferred by capillary action and UV-cross-linked to nylon membranes (Nytran SuPerCharge; Whatman [GE Healthcare Biosciences], Pittsburgh, PA, USA). Probes of PCR-amplified K3L, or an adjacent genomic region, were generated by biotin random prime labeling, hybridized, and bands detected with a North2South Chemiluminescent detection kit (Pierce [Thermo Scientific], Rockford, IL, USA).

For western blots, cells were infected with vaccinia for 48 hr and whole-cell lysates were collected in 2% SDS. Proteins were resolved by SDS-PAGE (12% Tris-glycine gel; Invitrogen, Carlsbad, CA, USA) and transferred to nitrocellulose membranes (Pierce). Proteins were detected with anti-K3L antibody (1:2,000; gift from J. Tartaglia), serum from rabbits infected with vaccinia virus (1:1,000; gift from B. Moss), and antiactin antibody (1:1,000; Abcam, Cambridge, UK).

siRNA Knockdown of K3L

Three double-stranded siRNAs were designed to target expression of K3L (1: 5′-aagauaaauguauagagcauaaucc(uu)-3′; 2: 5′-uucaacauaucuauccauaug cauc(uu)-3′; 3: 5′-uauaaucaacucuaaucacuuuaac(uu)-3′; IDTDNA). HeLa cells were transfected with one of the siRNA molecules (10 nM) or a nontargeting control siRNA that has at least four mismatches to all ORFs from several genomes (Dharmacon[Thermo Scientific]) using Lipofectamine 2000 (Invitrogen). Twenty-four hours after transfection, cells were infected with vaccinia for 48 hr. Expression of K3L was measured by western blot and β-gal activity was measured as a proxy for virus replication.

ACCESSION NUMBERS

Genomic sequence files are available from the NCBI Sequence Read Archive under accession SRP013416.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.05.049.

ACKNOWLEDGMENTS

We thank B. Jacobs for virus strains, J. Tartaglia for K3L antibody, B. Moss for vaccinia-reactive rabbit serum, and G. Smith for RecBCD and advice on recombination. We thank M. Daugherty, M. Dunham, M. Emerman, A. Frost, J. Kerns, M. Patel, and S. Sawyer for critical reading of the manuscript and B. Stackhouse for technical assistance with sequencing library preparation. This work is supported by the National Institutes of Health (NIH) K99/R00 award (GM090042 to N.C.E.), NIH grant (1R21CA160080-01 to J.S.), NIH grant (AI26672 to A.P.G.), a National Science Foundation (NSF) Graduate Research Fellowship (J.O.K.), and an NSF CAREER award (H.S.M.). H.S.M. is an Early Career Scientist of Howard Hughes Medical Institute. J.S. is a member of the scientific advisory board or serves as a consultant for Ariosa Diagnostics, Stratos Genomics, Good Start Genetics, and Adaptive TCR.

Received: April 30, 2012
Revised: May 29, 2012
Accepted: May 30, 2012
Published: August 16, 2012

REFERENCES

Adey, A., Morrison, H.G., Asan, Xun, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., and Shendure, J. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol. *11*, R119.

Afonso, C.L., Tulman, E.R., Lu, Z., Zsak, L., Kutish, G.F., and Rock, D.L. (2000). The genome of fowlpox virus. J. Virol. *74*, 3815–3831.

Andersson, D.I., Slechta, E.S., and Roth, J.R. (1998). Evidence that gene amplification underlies adaptive mutability of the bacterial lac operon. Science *282*, 1133–1135.

Bahar, M.W., Graham, S.C., Chen, R.A., Cooray, S., Smith, G.L., Stuart, D.I., and Grimes, J.M. (2011). How vaccinia virus has evolved to subvert the host immune response. J. Struct. Biol. *175*, 127–134.

Beattie, E., Denzler, K.L., Tartaglia, J., Perkus, M.E., Paoletti, E., and Jacobs, B.L. (1995). Reversal of the interferon-sensitive phenotype of a vaccinia virus lacking E3L by expression of the reovirus S4 gene. J. Virol. *69*, 499–505.

Beaumont, H.J., Gallie, J., Kost, C., Ferguson, G.C., and Rainey, P.B. (2009). Experimental evolution of bet hedging. Nature *462*, 90–93.

Bergthorsson, U., Andersson, D.I., and Roth, J.R. (2007). Ohno's dilemma: evolution of new genes under continuous selection. Proc. Natl. Acad. Sci. USA *104*, 17004–17009.

Bratke, K.A., and McLysaght, A. (2008). Identification of multiple independent horizontal gene transfers into poxviruses using a comparative genomics approach. BMC Evol. Biol. *8*, 67.

Cairns, J., Overbaugh, J., and Miller, S. (1988). The origin of mutants. Nature *335*, 142–145.

Cell

Crotty, S., Cameron, C.E., and Andino, R. (2001). RNA virus error catastrophe: direct molecular test by using ribavirin. Proc. Natl. Acad. Sci. USA *98*, 6895–6900.

Dar, A.C., and Sicheri, F. (2002). X-ray crystal structure and functional analysis of vaccinia virus K3L reveals molecular determinants for PKR subversion and substrate recognition. Mol. Cell *10*, 295–305.

Davies, M.V., Chang, H.W., Jacobs, B.L., and Kaufman, R.J. (1993). The E3L and K3L vaccinia virus gene products stimulate translation through inhibition of the double-stranded RNA-dependent protein kinase by different mechanisms. J. Virol. *67*, 1688–1692.

Dawkins, R., and Krebs, J.R. (1979). Arms races between and within species. Proc. R. Soc. Lond. B Biol. Sci. *205*, 489–511.

Demuth, J.P., and Hahn, M.W. (2009). The life and death of gene families. Bioessays *31*, 29–39.

Drake, J.W., and Holland, J.J. (1999). Mutation rates among RNA viruses. Proc. Natl. Acad. Sci. USA *96*, 13910–13913.

Drake, J.W., Charlesworth, B., Charlesworth, D., and Crow, J.F. (1998). Rates of spontaneous mutation. Genetics *148*, 1667–1686.

Drummond, J.T., Genschel, J., Wolf, E., and Modrich, P. (1997). DHFR/MSH3 amplification in methotrexate-resistant cells alters the hMutSalpha/hMutSbeta ratio and reduces the efficiency of base-base mismatch repair. Proc. Natl. Acad. Sci. USA *94*, 10144–10149.

Dunham, M.J., Badrane, H., Ferea, T., Adams, J., Brown, P.O., Rosenzweig, F., and Botstein, D. (2002). Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA *99*, 16144–16149.

Eichler, D.C., and Lehman, I.R. (1977). On the role of ATP in phosphodiester bond hydrolysis catalyzed by the recBC deoxyribonuclease of Escherichia coli. J. Biol. Chem. *252*, 499–503.

Elde, N.C., and Malik, H.S. (2009). The evolutionary conundrum of pathogen mimicry. Nat. Rev. Microbiol. *7*, 787–797.

Elde, N.C., Child, S.J., Geballe, A.P., and Malik, H.S. (2009). Protein kinase R reveals an evolutionary model for defeating viral mimicry. Nature *457*, 485–489.

Emerman, M., and Malik, H.S. (2010). Paleovirology—modern consequences of ancient viruses. PLoS Biol. *8*, e1000301.

Esposito, J., Condit, R., and Obijeski, J. (1981). The preparation of orthopoxvirus DNA. J. Virol. Methods *2*, 175–179.

Esposito, J.J., Sammons, S.A., Frace, A.M., Osborne, J.D., Olsen-Rasmussen, M., Zhang, M., Govil, D., Damon, I.K., Kline, R., Laker, M., et al. (2006). Genome sequence diversity and clues to the evolution of variola (smallpox) virus. Science *313*, 807–812.

Foster, P.L. (1994). Population dynamics of a Lac- strain of Escherichia coli during selection for lactose utilization. Genetics *138*, 253–261.

Goebel, S.J., Johnson, G.P., Perkus, M.E., Davis, S.W., Winslow, J.P., and Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. Virology *179*, 247–266, 517–263.

Hakki, M., and Geballe, A.P. (2005). Double-stranded RNA binding by human cytomegalovirus pTRS1. J. Virol. *79*, 7311–7318.

Harrison, S.C., Alberts, B., Ehrenfeld, E., Enquist, L., Fineberg, H., McKnight, S.L., Moss, B., O'Donnell, M., Ploegh, H., Schmid, S.L., et al. (2004). Discovery of antivirals against smallpox. Proc. Natl. Acad. Sci. USA *101*, 11178–11192.

Hastings, P.J., Bull, H.J., Klump, J.R., and Rosenberg, S.M. (2000). Adaptive amplification: an inducible chromosomal instability mechanism. Cell *103*, 723–731.

Hendrickson, H., Slechta, E.S., Bergthorsson, U., Andersson, D.I., and Roth, J.R. (2002). Amplification-mutagenesis: evidence that "directed" adaptive mutation and general hypermutability result from growth with a selected gene amplification. Proc. Natl. Acad. Sci. USA *99*, 2164–2169.

Hendrickson, R.C., Wang, C., Hatcher, E.L., and Lefkowitz, E.J. (2010). Orthopoxvirus genome evolution: the role of gene loss. Viruses *2*, 1933–1967.

Kaufman, R.J. (2000). Double-stranded RNA-activated protein kinase PKR. In Translational Control of Gene Expression, N. Sonenberg, J.W.B. Hershey, and M.B. Mathews, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 503–527.

Kawagishi-Kobayashi, M., Silverman, J.B., Ung, T.L., and Dever, T.E. (1997). Regulation of the protein kinase PKR by the vaccinia virus pseudosubstrate inhibitor K3L is dependent on residues conserved between the K3L protein and the PKR substrate eIF2alpha. Mol. Cell. Biol. *17*, 4146–4158.

Kerr, P.J., Hone, J., Perrin, L., French, N., and Williams, C.K. (2010). Molecular and serological analysis of the epidemiology of myxoma virus in rabbits. Vet. Microbiol. *143*, 167–178.

Kugelberg, E., Kofoid, E., Andersson, D.I., Lu, Y., Mellor, J., Roth, F.P., and Roth, J.R. (2010). The tandem inversion duplication in Salmonella enterica: selection drives unstable precursors to final mutation types. Genetics *185*, 65–80.

Langland, J.O., and Jacobs, B.L. (2002). The role of the PKR-inhibitory genes, E3L and K3L, in determining vaccinia virus host range. Virology *299*, 133–141.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, Y., Carroll, D.S., Gardner, S.N., Walsh, M.C., Vitalis, E.A., and Damon, I.K. (2007). On the origin of smallpox: correlating variola phylogenics with historical smallpox records. Proc. Natl. Acad. Sci. USA *104*, 15787–15792.

Lipinski, K.J., Farslow, J.C., Fitzpatrick, K.A., Lynch, M., Katju, V., and Bergthorsson, U. (2011). High spontaneous rate of gene duplication in Caenorhabditis elegans. Curr. Biol. *21*, 306–310.

Lynch, M. (2010). Evolution of the mutation rate. Trends Genet. *26*, 345–352.

Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C.R., Dopman, E.B., Dickinson, W.J., Okamoto, K., Kulkarni, S., Hartl, D.L., and Thomas, W.K. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc. Natl. Acad. Sci. USA *105*, 9272–9277.

McLysaght, A., Baldi, P.F., and Gaut, B.S. (2003). Extensive gene gain associated with adaptive evolution of poxviruses. Proc. Natl. Acad. Sci. USA *100*, 15655–15660.

Meisinger-Henschel, C., Schmidt, M., Lukassen, S., Linke, B., Krause, L., Konietzny, S., Goesmann, A., Howley, P., Chaplin, P., Suter, M., and Hausmann, J. (2007). Genomic sequence of chorioallantois vaccinia virus Ankara, the ancestor of modified vaccinia virus Ankara. J. Gen. Virol. *88*, 3249–3259.

Meyer, J.R., Dobias, D.T., Weitz, J.S., Barrick, J.E., Quick, R.T., and Lenski, R.E. (2012). Repeatability and contingency in the evolution of a key innovation in phage lambda. Science *335*, 428–432.

Meyerson, N.R., and Sawyer, S.L. (2011). Two-stepping through time: mammals and viruses. Trends Microbiol. *19*, 286–294.

Montville, R., Froissart, R., Remold, S.K., Tenaillon, O., and Turner, P.E. (2005). Evolution of mutational robustness in an RNA virus. PLoS Biol. *3*, e381.

Moss, B. (2007). Poxviridae: the viruses and their replication. In Fields Virology, D.M. Knipe and P.M. Howley, eds. (Philadelphia, PA: Lippincott Williams & Wilkins), pp. 2905–2946.

Moss, B., Winters, E., and Cooper, N. (1981). Instability and reiteration of DNA sequences within the vaccinia virus genome. Proc. Natl. Acad. Sci. USA *78*, 1614–1618.

Nair, S., Miller, B., Barends, M., Jaidee, A., Patel, J., Mayxay, M., Newton, P., Nosten, F., Ferdig, M.T., and Anderson, T.J. (2008). Adaptive copy number evolution in malaria parasites. PLoS Genet. *4*, e1000243.

Pfeiffer, J.K., and Kirkegaard, K. (2005). Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. PLoS Pathog. *1*, e11.

Prånting, M., and Andersson, D.I. (2011). Escape from growth restriction in small colony variants of Salmonella typhimurium by gene amplification and mutation. Mol. Microbiol. *79*, 305–315.

Rothenburg, S., Seo, E.J., Gibbs, J.S., Dever, T.E., and Dittmar, K. (2009). Rapid evolution of protein kinase PKR alters sensitivity to viral inhibitors. Nat. Struct. Mol. Biol. *16*, 63–70.

Sandegren, L., and Andersson, D.I. (2009). Bacterial gene amplification: implications for the evolution of antibiotic resistance. Nat. Rev. Microbiol. *7*, 578–588.

Schimke, R.T. (1986). Methotrexate resistance and gene amplification. Mechanisms and implications. Cancer *57*, 1912–1917.

Searles, R.P., Bergquam, E.P., Axthelm, M.K., and Wong, S.W. (1999). Sequence and genomic analysis of a Rhesus macaque rhadinovirus with similarity to Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8. J. Virol. *73*, 3040–3053.

Slabaugh, M.B., and Mathews, C.K. (1986). Hydroxyurea-resistant vaccinia virus: overproduction of ribonucleotide reductase. J. Virol. *60*, 506–514.

Slabaugh, M.B., Roseman, N.A., and Mathews, C.K. (1989). Amplification of the ribonucleotide reductase small subunit gene: analysis of novel joints and the mechanism of gene duplication in vaccinia virus. Nucleic Acids Res. *17*, 7073–7088.

Smith, G.L., and Moss, B. (1983). Infectious poxvirus vectors have capacity for at least 25 000 base pairs of foreign DNA. Gene *25*, 21–28.

Smith, G.R., Amundsen, S.K., Dabert, P., and Taylor, A.F. (1995). The initiation and control of homologous recombination in Escherichia coli. Philos. Trans. R. Soc. Lond. B Biol. Sci. *347*, 13–20.

Stambuk, B.U., Dunn, B., Alves, S.L., Jr., Duval, E.H., and Sherlock, G. (2009). Industrial fuel ethanol yeasts contain adaptive copy number changes in genes involved in vitamin B1 and B6 biosynthesis. Genome Res. *19*, 2271–2278.

Sun, S., Berg, O.G., Roth, J.R., and Andersson, D.I. (2009). Contribution of gene amplification to evolution of increased antibiotic resistance in Salmonella typhimurium. Genetics *182*, 1183–1195.

Van Valen, L. (1973). A new evolutionary law. Evol. Theor. *1*, 1–30.

Vignuzzi, M., Stone, J.K., and Andino, R. (2005). Ribavirin and lethal mutagenesis of poliovirus: molecular mechanisms, resistance and biological implications. Virus Res. *107*, 173–181.

Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E., and Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature *439*, 344–348.

Weber, F., Wagner, V., Rasmussen, S.B., Hartmann, R., and Paludan, S.R. (2006). Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. J. Virol. *80*, 5059–5064.

Werden, S.J., Rahman, M.M., and McFadden, G. (2008). Poxvirus host range genes. Adv. Virus Res. *71*, 135–171.

Wichman, H.A., Badgett, M.R., Scott, L.A., Boulianne, C.M., and Bull, J.J. (1999). Different trajectories of parallel evolution during viral adaptation. Science *285*, 422–424.

Supplemental Tables

Table S1. Genome Amplification Break Points in Replicates of Vaccinia Virus near K3L, Related to Figure 2

breakpoint	K4L break	K2L break	A*	B*	C*	parent***
1	31,725	30,296	1,757	2	3	0
2	31,573	28,284**	0	338	1	0
3	30,837	30,284	6	2,444	1,938	0
4	30,840	30,287	5	2,269	202	0
	total aligned reads:		4.5E+06	5.9E+06	3.7E+06	1.7E+06

break = breakpoint coordinate in reference genome (K2L ORF 29204..30313, K4L ORF 30681..31955)
* number of sequencing reads with breakpoint (multiplexed sequencing run)
** breakpoint occurs upstream of K2L ORF in K1L
*** parent strain was sequenced separately from the passage 10 replicates

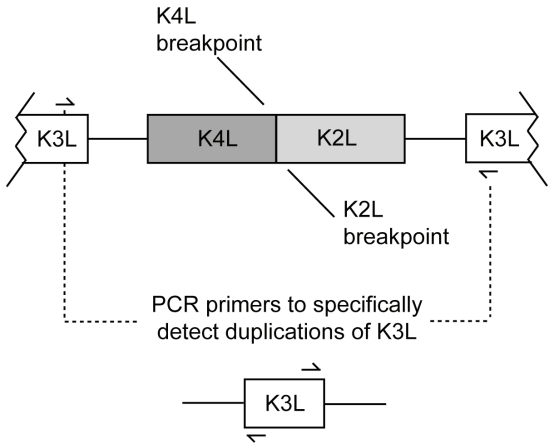


Table S2. Polymorphisms in Vaccinia Virus Populations from Experimental Evolution, Related to Figure 2

gene	variant	sub.	parent*	A*	B*	C*
K3L	H47R	T to C	0/917 (0%)	148/5510 (3%)	4/7627 -	541/4687 (12%)
E9L	E495G	T to C	0/825 (0%)	9/1818 -	11/2383 -	923/1475 (63%)
H2R	syn	C to T	0/631 (0%)	12/1693 -	905/2338 (39%)	11/1368 -

syn = synonymous substitution

* Showing a ratio of reads with the polymorphism to total reads. A 1% frequency cut-off was used to account for potential barcode misassignment in the multiplexed sequencing run with replicates A-C.

Table S3. Duplication-Specific Primers at Locations across Vaccinia Genomes, Related to Figure 5

ORF	Forward primer	Coord.*	Reverse primer	Coord.*
C20L**	GAT AGG TTC GAC ATT CAG AAA GAC G	6,874-6,850	GTT GAA GGT AAG TGT GTA ACC GC	7,093-7,115
C3L	CTC CAG GAA ATA CAT GGA AGC CG	21,806-21,784	CGC ATC CTA TTC CCA ACA ATG TC	22,503-22,525
K3L	GGG ATA AAC TGG TAG GGA AAA CTG TAA AAG	30,466-30,437	CAG AGT GAG GAT AGT CAA AAA GAT AAA TGT ATA G	30,521-30,554
F13L	GAG ATT GGG TAT CTA GCC ACA GTA	44,001-43,978	TTC TAC CAG CCT ACA TTT TGC TCC	45,029-45,052
I7L	CGT ATG GAA AAA TGG ACC AAA AAG TC	68,886-68,861	CAT CTA TAT TGC TAC ATA ATC CAG CTA GTG	69,963-69,992
J2R	GAT ATG TAT CAA TCG GTG TGT AGA AAG TG	84,344-84,372	CTG TAC TTT TAC CTG AAA ACA TGG GG	83,912-83,887
D5R	GCC AAT GTT AGT GAT GAA TAT CTG CA	103,718-103,743	CCT CGA TAA TAA AAT CTT GAA TAG CCG T	101,696-101,669
A4L	GGG CAA TAG ATC AAA TTC TAC AAC TAG T	116,480-116,453	GGT GTA GAG GAT TCT GCC AGC	117,221-117,241
A29L	AAG CAT TGT CTG ATG CGT GTA A	139,824-139,803	GTT CGA GTT CAA CGA CGA TTG A	140,644-140,665
A52R	GGA GAA ATG AGA AAC TGT TTT CTA GAT GG	158,199-158,227	TTT CCC TCC TAG TAG TCA CCG	157,820-157,800
B20R	GAA ACA CAA GAT CAA AGT ACA CAA CAC A	179,645-179,672	TCG GTG AGA TAC AAA TAC CTA GAT AGT C	179,343-179,316

* 'L' ORFs coordiate numbers based on complement sequence, Copenhagen strain
** same for B26R (coordinates 184,864-184,888; 184,645-184,623)

Table S4. Sequenced Duplications in Vaccinia Strains, Related to Figure 5

ORF (strain)	'right' break*	'left' break*	Fully duplicated ORFs
C20L (WR)	7,345	6,659	C20L
C3L (WR)	22,675	21,741	C3L
K3L (WR)	30,738	30,261	K3L
K3L (WR)	30,587	30,399	-
B20R (VC2)	179,726	179,121	B20R
B20R (ΔE3L)	179,905	179,025	B20R
B26R** (WR)	185,174	184,579	B26R

* genomic coordinates from vaccinia Copenhagen reference (Genbank M35027)
** same as C20L region in reverse orientation

Table S5. Nonrecurring Genomic Break Points Consistent with Duplications, Related to Figure 5

virus strain	'right' break*	'left' break*	inferred duplicated ORFs
ΔE3L parent	156,405	143,153	A35R - A49R
ΔE3L parent	162,444	22,338	large genomic block
A	31,413	30,525	-**
A	67,627	66,482	-***
A	104,930	34,700	large genomic block
A	107,526	107,330	-
B	29,837	22,298	C1L, N1L, N2L, M1L, M2L, K1L
B	47,387	24,975	N2L, M1L, M2L, K1L-K7L, F1L-F17L**
B	47,467	25,066	N2L, M1L, M2L, K1L-K7L, F1L-F17L**
B	55,156	55,104	-
B	141,790	141,697	-
B	169,683	168,828	B7R
C	24,505	23,524	-
C	90,475	87,550	'H orf A'***
C	104,395	104,298	-
C	167,515	89,039	large genomic block
C	171,274	100,726	large genomic block

* genomic coordinates from vaccinia Copenhagen reference (Genbank M35027)
** includes K3L
***confirmed by PCR and sequencing

Table S6. Vaccinia Virus Genome Comparison between ΔE3L and Copenhagen Reference Strain (Genbank M35027), Related to Figure 2

Coordinate (bp)	Copenhagen nucleotide	ΔE3L nucleotide	Variant freq (%)	Gene	Variant
16,033	A	G	98.6	intergenic	-
23,443	C	T	99.8	C2L	syn
24,256	G	C	100.0	C1L	syn
25,525	G	C	100.0	N2L	P121R
35,080-1	CG	GC	99.3	F3L	T313S
44,312	A	G	99.6	F13L	syn
46,742-3	CG	GC	99.2	F16L	R10A
77,258	A	T	99.9	G7L	V348E
81,834	G	A	99.6	L3L	L1384F
85,139	T	C	99.8	J3R	V209A
104,656	C	T	99.5	D6R	T180M
114,376	-	C	88.4	intergenic	-
145,839	T	A	97.2	intergenic	-
148,904	A	T	99.7	A41L	syn
152,699	-	C	81.2	A46R	frameshift

syn = synonymous substitution

CHAPTER 3

EMERGENCE OF A VIRAL RNA POLYMERASE VARIANT DURING GENE COPY NUMBER AMPLIFICATION PROMOTES RAPID EVOLUTION OF VACCINIA VIRUS

Reprinted with permission from Cone, K.R., Kronenberg, Z.N., Yandell, M., and Elde, N.C. (2017). Emergence of a viral RNA polymerase variant during gene copy number amplification promotes rapid evolution of vaccinia virus. J. Virol. *91*. e01428-16.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

GENETIC DIVERSITY AND EVOLUTION

Emergence of a Viral RNA Polymerase Variant during Gene Copy Number Amplification Promotes Rapid Evolution of Vaccinia Virus

Kelsey R. Cone,[a] Zev N. Kronenberg,[a,b]* Mark Yandell,[a,b] Nels C. Elde[a]

Department of Human Genetics, University of Utah, Salt Lake City, Utah, USA[a]; Utah Center for Genetic Discovery, University of Utah, Salt Lake City, Utah, USA[b]

ABSTRACT Viruses are under relentless selective pressure from host immune defenses. To study how poxviruses adapt to innate immune detection pathways, we performed serial vaccinia virus infections in primary human cells. Independent courses of experimental evolution with a recombinant strain lacking E3L revealed several high-frequency point mutations in conserved poxvirus genes, suggesting important roles for essential poxvirus proteins in innate immune subversion. Two distinct mutations were identified in the viral RNA polymerase gene A24R, which seem to act through different mechanisms to increase virus replication. Specifically, a Leu18Phe substitution encoded within A24R conferred fitness trade-offs, including increased activation of the antiviral factor protein kinase R (PKR). Intriguingly, this A24R variant underwent a drastic selective sweep during passaging, despite enhanced PKR activity. We showed that the sweep of this variant could be accelerated by the presence of copy number variation (CNV) at the K3L locus, which in multiple copies strongly reduced PKR activation. Therefore, adaptive cases of CNV can facilitate the accumulation of point mutations separate from the expanded locus. This study reveals how rapid bouts of gene copy number amplification during accrual of distant point mutations can potently facilitate poxvirus adaptation to host defenses.

IMPORTANCE Viruses can evolve quickly to defeat host immune functions. For poxviruses, little is known about how multiple adaptive mutations emerge in populations at the same time. In this study, we uncovered a means of vaccinia virus adaptation involving the accumulation of distinct genetic variants within a single population. We identified adaptive point mutations in the viral RNA polymerase gene A24R and, surprisingly, found that one of these mutations activates the nucleic acid sensing factor PKR. We also found that gene copy number variation (CNV) can provide dual benefits to evolving virus populations, including evidence that CNV facilitates the accumulation of a point mutation distant from the expanded locus. Our data suggest that transient CNV can accelerate the fixation of mutations conferring modest benefits, or even fitness trade-offs, and highlight how structural variation might aid poxvirus adaptation through both direct and indirect actions.

KEYWORDS RNA polymerase, experimental evolution, genome analysis, innate immunity, poxvirus, vaccinia virus

Received 19 July 2016 **Accepted** 29 November 2016

Accepted manuscript posted online 7 December 2016

Citation Cone KR, Kronenberg ZN, Yandell M, Elde NC. 2017. Emergence of a viral RNA polymerase variant during gene copy number amplification promotes rapid evolution of vaccinia virus. J Virol 91:e01428-16. https://doi.org/10.1128/JVI.01428-16.

Editor Grant McFadden, The Biodesign Institute, Arizona State University

Address correspondence to Nels C. Elde, nelde@genetics.utah.edu.

* Present address: Zev N. Kronenberg, Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA.

Although the mutation rates of animal viruses are much higher than those of their hosts, the point mutation rate varies greatly between different types of viruses (1–4). For example, some double-stranded DNA (dsDNA) viruses have point mutation rates that are orders of magnitude lower than those of RNA viruses (3). Poxviruses, for instance, are predicted to have relatively low point mutation rates due to 3′-5′ proofreading activity of the viral DNA polymerase (3, 5, 6). While recent estimates

suggest a higher point mutation rate for poxviruses than for other dsDNA viruses (7, 8), these rates are still lower than those for most RNA viruses. Observations that viruses with various mutation rates flourish in shared hosts strongly predict that successful adaptation of dsDNA viruses, including poxviruses, relies on mechanisms in addition to the rapid sampling of point mutations.

Vaccinia virus (VACV) provides a useful model system for studying poxvirus evolution due to the vast repertoire of available molecular tools and numerous well-characterized interactions between VACV-encoded factors and the host innate immune system. One key interface involves interactions between VACV and the host nucleic acid sensor protein kinase R (PKR). Upon binding viral dsRNA, PKR phosphorylates the eukaryotic translation initiation factor eIF2α. Phosphorylation of eIF2α leads to a severe block in protein translation and to attenuated viral replication. Like many poxviruses, VACV encodes E3L and K3L, which inhibit PKR via different mechanisms (9–11).

Consistent with their role as key host range factors, VACV E3L and K3L vary in the ability to block PKRs from different host species (9, 12). In particular, K3L is a poor inhibitor of human PKR, such that a VACV strain lacking E3L (ΔE3L) (13) exhibits a severe replication defect during infection of HeLa cells (12). This deficiency places strong selective pressure on the virus to adapt to counteract PKR. Previous courses of experimental evolution of ΔE3L in HeLa cells revealed a recombination-based "genomic accordion" mechanism, in which copy number variation (CNV) of the K3L gene allowed rapid adaptation by inhibiting the activity of PKR (14). Although whole-genome sequencing of VACV revealed that each of the three replicate populations in this experiment acquired an increased K3L copy number, there were differences between populations in the recombination breakpoints as well as unique high-frequency point mutations throughout the genome. These differences suggest that experimental evolution is just beginning to uncover adaptive mutations and mechanisms contributing to poxvirus evolution. Further analysis of evolving populations under different conditions might reveal unrecognized means of virus adaptation.

In this study, we identified adaptive mutations in VACV genomes that arose during serial infections of primary human fibroblast (HF) cells. Interestingly, two nonsynonymous point mutations, from independent replicate populations, were identified within the A24R gene, which encodes a catalytic subunit of the viral RNA polymerase (vRNAP). Experimental rescue analysis indicates that either a Leu18Phe or Lys452Asn amino acid change in the A24R gene product is sufficient to provide a replication gain to the ΔE3L virus in HF cells. Additionally, the A24R mutations we found seem to act through mechanisms distinct from those of previously identified adaptive A24R mutations to improve viral fitness. We also show that by blocking PKR activation, the K3L CNV arising in our virus populations enhanced the accumulation of a point mutation in A24R. This work provides a new view of how the rapid fixation of a beneficial point mutation, often described as a selective sweep, can occur concurrently with recombination-mediated adaptation in a viral population, and it illuminates a fundamental mechanism for how structural variants (SVs) might enhance poxvirus adaptation.

RESULTS

Replication gains following serial infection of primary HF cells. To study mechanisms of viral adaptation, we performed serial VACV infections of primary human fibroblast (HF) cells. ΔE3L replicates poorly in HeLa cells (12), so we first tested whether the defect is similar for infection of primary human fibroblasts. While both ΔE3L and the parental wild-type Copenhagen strain of VACV (VC-2) replicated equally in a permissive hamster (BHK) cell line during 48-h infections, ΔE3L displayed an even stronger growth defect compared to the growth of VC-2 in HF cells than in HeLa cells (Fig. 1A). This result is consistent with cell type-specific innate immune responses to viral infection placing different selective pressures on the virus (15), even between different human cell lines.

To test how VACV might adapt to primary human cells, we performed serial infections of HF cells with ΔE3L in triplicate. Using viral titer as a measure of fitness, we observed rapid gains in fitness over the course of 10 passages for all three replicate

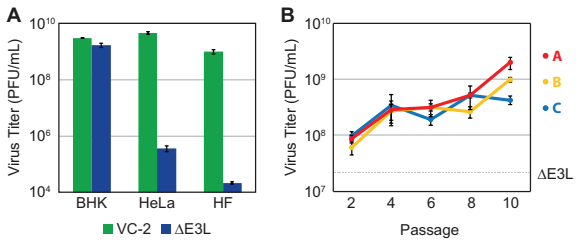


FIG 1 Rapid adaptation of ΔE3L during experimental evolution in HF cells. (A) Cells were infected with wild-type VC-2 or ΔE3L virus (MOI = 0.1) for 48 h. (B) Triplicate populations of ΔE3L virus were passaged 10 times in HF cells. Equal volumes of virus from every other passage were expanded in BHK cells for 48 h and titrated simultaneously. The similarly expanded parental ΔE3L virus titer is indicated by the dotted line. All viral titers were measured in BHK cells by 48-h plaque assays performed in triplicate, and data are mean PFU per milliliter ± standard deviations.

virus populations (Fig. 1B). Despite modest replication of the parental ΔE3L virus in HF cells compared to HeLa cells, we observed comparable gains in replication of ΔE3L virus populations over the course of our infections of HF cells (Fig. 1B) and HeLa cells, as previously reported (14).

High-frequency point mutations in evolved virus populations. To define genetic changes that might account for increases in viral fitness, we used deep sequencing of viral genomes from each of the replicate populations after 10 rounds of serial infection (P10). We obtained an average of >2,000× coverage across the genome for each P10 population. Excluding the inverted terminal repeat regions (1 to 5,000 bp and 186,737 to 191,737 bp), we identified nine single nucleotide polymorphisms (SNPs) not present in the ΔE3L parent virus at a frequency of >1% for any P10 population, in addition to 20 shared differences compared to the VC-2 reference strain (Table 1). The nine SNPs present in at least one of the three P10 populations but not in the ΔE3L parent virus represent potentially adaptive mutations. All nine SNPs lie within open reading frames, and notably, seven of them result in nonsynonymous amino acid changes or frameshifts (Fig. 2A). Remarkably, the highest-frequency SNP in each population caused a substitution in an essential gene conserved among poxviruses, resulting in a Leu18Phe amino acid substitution in A24R, encoding a catalytic subunit of the viral RNA polymerase (vRNAP) (16); a Glu495Gly amino acid substitution in E9L, encoding the viral DNA polymerase (17, 18); and a Trp44Cys amino acid substitution in F10L, encoding a kinase required for virion morphogenesis (19). These core genes are all involved in replication and assembly, suggesting a common adaptive advantage through altered replication cycle kinetics under nonoptimal conditions in which nucleic acid sensors are activated. This may represent an indirect selection for altered replication and assembly in addition to the direct influence of innate signaling pathways. Furthermore, another A24R mutation, resulting in a Lys452Asn amino acid substitution, was identified in a replicate population independently of the Leu18Phe variant. This result suggests that the A24R gene may be a common target for beneficial mutations in VACV, consistent with previous reports of other adaptive A24R mutations (20–22). Together, these high-frequency mutations suggest a common role for poxvirus genes encoding essential viral functions in adaptation to activated host innate immune responses and an altered cellular environment.

SVs in evolved virus populations. In addition to point mutations, genetic changes in the form of gene copy number variation (CNV) have previously been shown to play an adaptive role during poxvirus adaptation (14). To identify potentially adaptive structural variants (SVs) in our evolved virus populations, we analyzed the P10 sequences by using the SV analysis implemented in the program Wham (23). We found seven SVs with >10 reads to define both the 5′ and 3′ locations of recombination

TABLE 1 Point mutations in virus populations relative to the VC-2 reference sequence

		Nucleotide(s)		Allele frequency[b]			
Position	ORF[a]	Reference	Variant	ΔE3L	A P10	B P10	C P10
16033		A	G	1.00	1.00	1.00	1.00
23443		C	T	1.00	1.00	1.00	1.00
24256		G	C	1.00	1.00	1.00	1.00
25525		G	C	1.00	1.00	1.00	1.00
30295[c]	**K2L**	**G**	**A**	**0.00**	**0.14**	**0.00**	**0.02**
30752	**K4L**	**GT**	**G**	**0.00**	**0.00**	**0.00**	**0.11**
35080		C	G	0.85	0.69	0.67	0.63
35081		G	C	0.86	0.80	0.78	0.75
40751	**F10L**	**C**	**A**	**0.00**	**0.00**	**0.00**	**0.60**
44312		A	G	1.00	1.00	1.00	1.00
46730[c]		T	A	0.07	0.17	0.18	0.13
46742		C	G	1.00	1.00	1.00	1.00
46743		G	C	1.00	1.00	1.00	1.00
51482[c]		A	C	1.00	0.96	1.00	1.00
58304	**E9L**	**T**	**C**	**0.00**	**0.00**	**0.80**	**0.00**
69922		C	T	0.09	0.06	0.08	0.04
77258		A	T	1.00	1.00	1.00	1.00
81834		G	A	1.00	1.00	1.00	1.00
85139		T	C	1.00	1.00	1.00	1.00
95128		T	C	0.32	0.38	0.37	0.40
104656		C	T	1.00	1.00	1.00	1.00
120212	**A7L**	**G**	**T**	**0.00**	**0.00**	**0.31**	**0.00**
134369	**A24R**	**G**	**T**	**0.00**	**0.91**	**0.00**	**0.00**
135671	**A24R**	**G**	**T**	**0.00**	**0.00**	**0.00**	**0.14**
145840		T	A	1.00	1.00	1.00	1.00
148905		A	T	1.00	1.00	1.00	1.00
152700		G	GC	1.00	1.00	1.00	1.00
172688	**B13R**	**G**	**T**	**0.00**	**0.00**	**0.00**	**0.18**
175621	**B17L**	**CT**	**C**	**0.00**	**0.00**	**0.25**	**0.00**

[a]The ORF containing the mutation is shown for variant alleles (bold) not present in the ΔE3L parent virus (also see Fig. 2A).
[b]Alleles with frequencies of >0.01 are shown.
[c]Predicted false-positive variant stemming from a misaligned read at a structural variant breakpoint (see Table 2).

breakpoints in the virus populations (Fig. 2B; Table 2). Sanger sequencing corroborated each of these SVs to within one nucleotide of the putative breakpoint. The single SV identified in the ΔE3L parent virus was also present in all three P10 populations (breakpoint 7 in Table 2). This variant corresponds to the 11K vaccinia virus promoter introduced during the deletion of E3L (13), resulting in two copies of this sequence, found at different genomic locations (Fig. 2B). The remaining 6 SVs were present only in the P10 virus populations, and each of them was associated with the K3L locus. There was a corresponding increase in sequencing depth surrounding K3L for each P10 virus population, but this was absent in the parent population (Fig. 2B). Three of the SNPs from variant calling were located 1 bp from a structural variant breakpoint, and we therefore categorized these as false-positive calls due to misaligned reads (Table 1). This finding illustrates that SVs in poxvirus genomes should be considered when using standard variant calling methods to identify SNPs.

A similar copy number amplification of the K3L locus was observed following serial infections of HeLa cells with ΔE3L (14). In the previous study, we showed that CNV at the K3L locus corresponds to an increased amount of K3L protein, which impairs PKR activation. Interestingly, two of the seven breakpoints identified in this study were identical to those observed in virus populations passaged in HeLa cells (14), suggesting the presence of these variants at a level below the limit of detection in the ΔE3L parent virus or indicating that these are preferential sites for recombination. These results, combined with studies demonstrating CNV in other poxvirus genes in response to selective pressure (24–26), continue to reveal CNV as a mechanism for rapid adaptation of VACV.

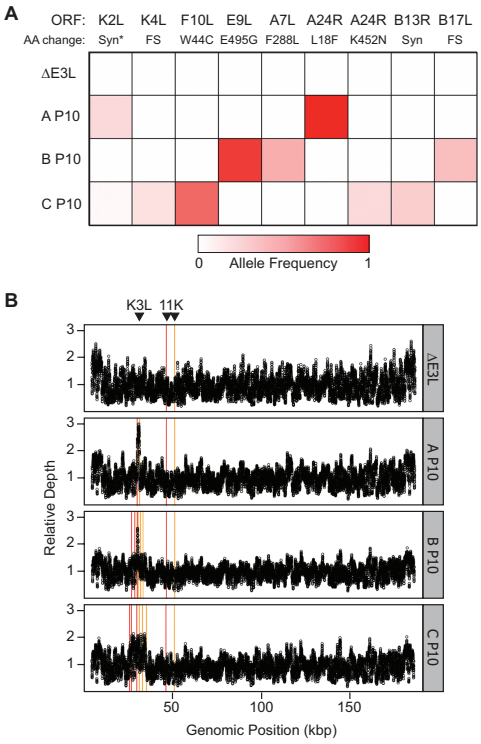


FIG 2 Genetic changes in virus populations following experimental evolution. (A) Allele frequencies were obtained by deep sequencing of viral genomic DNA from passage 10 virus populations compared to the ΔE3L parent virus. Alleles with a frequency of ≥0.01 that were not present in the parent virus are shown. The VACV open reading frame (ORF) and resulting amino acid change are listed at the top. FS, frameshift; Syn, synonymous. *, the K2L allele is likely a miscalled structural variant. (B) Relative depths of coverage across the viral genome are shown in black, excluding the inverted terminal repeat regions. Breakpoint positions of structural variants called by Wham are shown as vertical lines, with 5′ positions in red and 3′ positions in orange. The genomic locations of the K3L gene and the duplicated 11K promoter are indicated at the top.

A24R mutations increase viral fitness through distinct mechanisms. Following 10 passages in HF cells, each population of viruses harbored a unique set of mutations that might contribute to increased fitness. The most drastic example was a Leu18Phe amino acid substitution encoded within A24R that was nearly fixed (frequency = 0.91) in one virus population after 10 passages (Fig. 2A). The lack of other nonsynonymous point mutations in this replicate population above a frequency of 1% strongly suggests that this single mutation contributed to the observed increase in fitness (Fig. 1B). To test this hypothesis, we generated a recombinant virus with this A24R mutation in the parental ΔE3L strain (A24R^{Leu18Phe}). Growth curve analysis with the permissive BHK cell line revealed that the A24R^{Leu18Phe} and ΔE3L viruses replicated to similar titers (Fig. 3A). However, in HF cells, the A24R^{Leu18Phe} virus exhibited a significant increase in titer relative to that of ΔE3L from 48 to 72 h postinfection. Thus, the single nucleotide change in A24R in the ΔE3L genetic background was sufficient to enhance viral fitness under selective pressure in HF cells.

To examine how a single amino acid substitution encoded within A24R might increase viral fitness, we mapped A24R mutations onto solved structures of RNA

TABLE 2 Structural variants in virus populations relative to the VC-2 reference sequence

Breakpoint	Genomic position[a]		Read support (no. of reads [5′, 3′])			
	5′	3′	ΔE3L	A P10	B P10	C P10
1	26302	35775	0, 0	0, 0	0, 0	112, 102
2	27105	32160	0, 0	0, 0	74, 70	0, 0
3	27322	33524	0, 0	0, 0	0, 0	105, 73
4	28875	33568	0, 0	0, 0	114, 74	0, 0
5	30287	30840	0, 0	0, 0	164, 125	0, 0
6	30296	31725	0, 0	190, 137	0, 0	96, 85
7	46731	51483	179, 260	159, 465	179, 472	200, 389

[a]Positions supported by >10 reads are shown.

polymerases. Previous work used the crystal structure of *Saccharomyces cerevisiae* RNA polymerase II to map A24R mutations encoding amino acid substitutions onto the homologous subunit in yeast, RBP2 (22). RBP2 is the second largest subunit of yeast RNA polymerase II and forms an active site of the enzyme with RBP1 (27). Using a similar approach, we generated an amino acid alignment between the VACV A24R gene product and *S. cerevisiae* RBP2 to predict the locations of amino acid substitutions encoded within A24R mutants on the *S. cerevisiae* RNA polymerase II crystal structure (PDB entry 1I50) (28). This analysis suggested that the Leu18Phe substitution encoded by the corresponding A24R mutant is located on a solvent-exposed surface of the polymerase distal from the active site of the enzyme (Fig. 3B). This surface might be involved in binding other factors to the polymerase, such that a Leu18Phe substitution may alter a protein-protein interaction(s). However, little is currently known about viral or host proteins that bind to the A24R gene product, making it difficult to predict the functional consequences of this mutation.

Distinct adaptive mutations in A24R demonstrate how vRNAP variation can affect viral fitness. Two A24R mutations were shown to influence transcription elongation, resulting in the production of short virus transcripts in response to isatin-β-thiosemicarbazone (IBT) selection (21, 22). IBT-resistant mutations have also been shown to reduce the activation of the host dsRNA sensor oligoadenylate synthetase (OAS) (29, 30). These studies suggest that changes to transcript length, dictated by vRNAP, can influence the activation of innate immune dsRNA sensors in the cell. Indeed, another A24R mutation was recently identified that reduces the activation of the dsRNA sensor PKR (20). Without E3L, viruses are more vulnerable to PKR and OAS activities, and thus we hypothesized that the A24R substitutions we identified might act through a similar mechanism to reduce the activation of dsRNA sensors.

To test whether either of the A24R variants from our evolved virus populations is sufficient to alter innate immune responses, we first determined whether dsRNA production was affected by the vRNAP variants. A dot blot of dsRNAs from infected cells revealed less total dsRNA in cells infected with either the A24R^{Leu18Phe} or A24R^{Lys452Asn} recombinant virus than that in cells infected with the ΔE3L virus (Fig. 3C). Viruses with the Leu18Phe substitution produced a notable reduction in dsRNA, which might reflect changes to transcript length or stability, or even changes to a subset of transcripts. Additionally, the overall reduction in dsRNA might have affected the activation of dsRNA sensors and contributed to the fitness increases observed for viruses harboring the A24R variants.

To test whether a reduction in dsRNA affects downstream nucleic acid sensing pathways, we measured the activities of the OAS/RNase L and PKR pathways in cells infected with recombinant viruses harboring the A24R mutant-encoded Leu18Phe or Lys452Asn substitution. We did not detect any notable difference in RNase L activity as judged by rRNA degradation following infection with either the A24R^{Leu18Phe} or A24R^{Lys452Asn} recombinant virus (Fig. 3D). This result suggests that the OAS/RNase L pathway was not significantly affected by these A24R substitutions, despite the reduction in dsRNA. We next tested whether either of the A24R variants reduced PKR

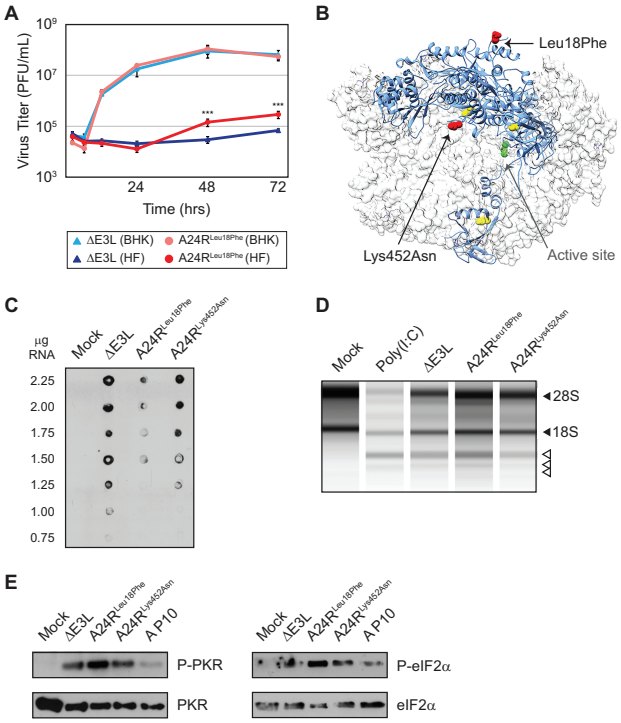


FIG 3 A24R mutations increase fitness through distinct mechanisms. (A) Single-step growth curve analysis (MOI = 5.0) was performed in triplicate with either the ΔE3L or A24R^{Leu18Phe} recombinant virus in BHK or HF cells. Viral titers were measured in BHK cells by a 72-h plaque assay, and data are mean PFU per milliliter ± standard deviations. ***, $P < 0.005$ relative to ΔE3L within each cell type, by 2-way analysis of variance (ANOVA) followed by Bonferroni's multiple-comparison test. (B) Positions of A24R mutations identified in this study (red spheres) and previously published data (yellow spheres) were mapped onto the homologous *S. cerevisiae* RNAP II structure (PDB entry 1I50). The RBP2 protein (homologous to the A24R gene product) is shown as blue ribbons in the context of the multisubunit polymerase shown in gray (the active site is shown in green). (C) dsRNA dot blot with decreasing amounts of total RNA from HF cells left untreated (mock) or infected with the ΔE3L, A24R^{Leu18Phe}, or A24R^{Lys452Asn} virus (MOI = 10.0) for 13 h. An image representative of three independent blots is shown. (D) rRNA degradation in HF cells left untreated (mock), transfected with poly(I·C), or infected with the ΔE3L, A24R^{Leu18Phe}, or A24R^{Lys452Asn} virus (MOI = 5.0) for 6 h. Filled arrowheads indicate 28S and 18S rRNAs, and open arrowheads indicate degradation products. (E) Immunoblot of phosphorylated or total PKR and phosphorylated or total eIF2α in HF cells infected with the ΔE3L, A24R^{Leu18Phe}, A24R^{Lys452Asn}, or replicate A passage 10 virus (MOI = 5.0) for 6 h. An image representative of five independent blots is shown.

activation as measured by changes in phosphorylated PKR and phosphorylated eIF2α protein levels in infected HF cells. Counter to our expectation, immunoblot analysis indicated that HF cells infected with the A24R^{Leu18Phe} virus repeatedly showed increased levels of both phosphorylated PKR and eIF2α (Fig. 3E). While the A24R^{Leu18Phe} virus conferred a replication benefit at late time points that may account for an increase in PKR activation, this is unlikely given the equivalent viral titers between the A24R^{Leu18Phe} and ΔE3L viruses at the 6-h time point (Fig. 3A) when we harvested total protein. In contrast, there were no substantial changes in PKR activation upon infection with the A24R^{Lys452Asn} virus compared to infection with ΔE3L. These data suggest that the A24R variants we identified work through mechanisms distinct from those for other known A24R substitutions to enhance viral fitness. Moreover, the increase in PKR

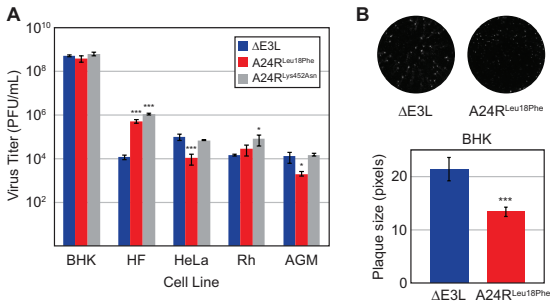


FIG 4 The A24R Leu18Phe variant confers fitness trade-offs. (A) Cell lines were infected in triplicate with the ΔE3L, A24R^{Leu18Phe}, or A24R^{Lys452Asn} virus (MOI = 0.1) for 48 h. Viral titers were measured in BHK cells by a 72-h plaque assay, and data are mean PFU per milliliter ± standard deviations. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.005$ relative to ΔE3L within each cell type, by one-way ANOVA followed by Dunnett's multiple-comparison test. Rh, rhesus macaque fibroblasts; AGM, African green monkey fibroblasts. (B) Average plaque size ± standard deviation for three independent wells of BHK cells infected with the ΔE3L or A24R^{Leu18Phe} virus (MOI = 0.1) for 48 h. A representative image is shown. ***, $P < 0.005$ by 2-tailed *t* test.

activation with the A24R^{Leu18Phe} virus was paradoxical, considering the reduction in dsRNA and the replication advantage we observed in HF cells.

The A24R Leu18Phe variant confers fitness trade-offs. Given that E3L and K3L act as host range factors, blocking innate immune activation in some hosts but not others (12), we tested the impact of A24R variation during infections of cells from other primate species. We performed triplicate 48-h infections in two human and two Old World primate cell lines: HF cells, HeLa cells, rhesus macaque fibroblasts, and African green monkey fibroblasts. In each of the four cell lines tested, the parental ΔE3L virus exhibited considerably reduced replication compared to that in a permissive BHK cell line (Fig. 4A). The A24R^{Lys452Asn} virus replicated equally to or better than ΔE3L in all primate cells tested, suggesting that the fitness increase provided by the A24R Lys452Asn variant is not species or cell type specific. In contrast, while A24R^{Leu18Phe} virus replication showed a 43-fold increase in HF cells relative to replication of the ΔE3L virus, titers were significantly reduced in HeLa cells and African green monkey fibroblasts. The clear differences in A24R^{Leu18Phe} viral titers between cells from two Old World monkey species and also between two human cell lines reveal a potential fitness trade-off for viruses harboring this substitution.

Consistent with a fitness trade-off in some cell types, the A24R^{Leu18Phe} virus displayed a small-plaque phenotype in permissive BHK cells (Fig. 4B). This phenotype is associated with defects in cell-to-cell spread (31–33), although the A24R^{Leu18Phe} virus produced approximately the same number of infectious particles as ΔE3L in BHK cells (Fig. 3A). The small-plaque phenotype in conjunction with differential replication among cell lines supports the idea that the A24R Leu18Phe variant is beneficial only under certain conditions. Furthermore, the growth benefit in HF cells and the growth defect in HeLa cells highlight that specific, species-independent differences between the cell lines can contribute to the success or failure of the A24R^{Leu18Phe} virus.

Accelerated sweep of the A24R Leu18Phe variant during adaptive gene copy number amplification. The A24R Leu18Phe variant rose to near fixation in a viral population during experimental evolution of ΔE3L in HF cells (Fig. 2; Table 1). The rapid dominance of the variant in this population was unusual given the slower accumulation dynamics of mutations within other replicate populations. For example, the A24R Lys452Asn variant provided a similar replication increase alone (Fig. 4A) yet reached a frequency of only 14%, in contrast to the 91% frequency of the A24R Leu18Phe variant (Fig. 5A). Furthermore, we might have predicted that the increased PKR activation

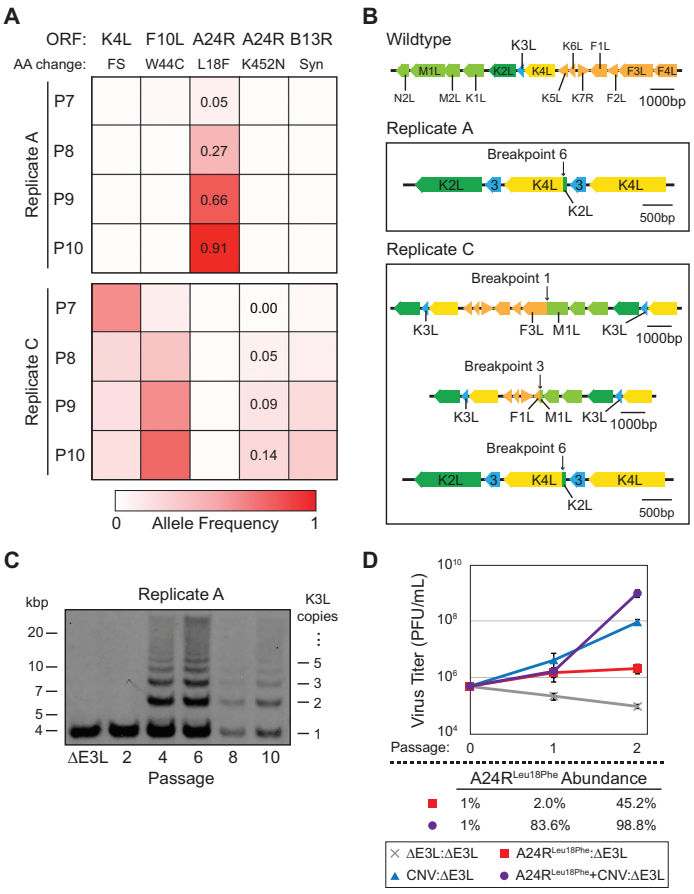


FIG 5 Copy number variation enhances the sweep of a point mutation. (A) Allele frequencies of >0.01 for replicates A and C were obtained by deep sequencing of P7 to P10 virus populations, and positions in ORFs and resulting amino acid changes are listed at the top, as in Fig. 2A. FS, frameshift; Syn, synonymous. (B) Genome structures from direct sequencing across the CNV breakpoints identified by Wham (Table 2) for replicate A and C passage 10 populations. (C) Southern blotting using a K3L-specific probe was performed on digested viral DNAs from the indicated viral populations. Sizes (left) and numbers of K3L copies (right) are shown. (D) Viruses were mixed as listed, at a ratio of 1:100, and passaged twice for 48 h in HF cells (MOI = 0.1). Titers were measured in BHK cells by 72-h plaque assays performed in triplicate, and data are mean PFU per milliliter ± standard deviations. A24R Leu18Phe variant abundances are listed below the graph, as input (P0) or as measured by deep sequencing (P1 and P2).

induced by the A24R Leu18Phe variant (Fig. 3E) would prevent the mutation from reaching fixation so rapidly. To more carefully determine how this mutation arose, we analyzed earlier passages during the course of experimental evolution. Remarkably, the A24R Leu18Phe variant was detectable only starting at passage 8 as judged by sequencing of A24R amplicons in each passage, prompting us to deep sequence virus populations from passages 7 to 9. Genome sequence analysis revealed that the mutation rapidly increased in frequency with each successive passage late in the experiment (Fig. 5A). Consistent with our earlier analysis, this was the only verified SNP

present at frequencies above 1% in passages 7 to 10 (replicate A), which is in stark contrast to the findings for replicate C, in which multiple mutations across the genome fluctuated in frequency from passages 7 to 10 (Fig. 5A). The rapid accumulation of the A24R Leu18Phe variant, as well as a lack of any other mutations across the genome, is consistent with a strong selective sweep of the mutation in the replicate A population. However, because the A24R^{Leu18Phe} virus also increased PKR activation, the question remained as to how the mutation induces increased virus replication.

The increased copy number of the K3L gene in the P10 population provided a clue to how the A24R mutation might rapidly sweep to fixation despite increasing PKR activation. There was only one recombination breakpoint at P10 for replicate A, as opposed to multiple breakpoints for replicates B and C (Fig. 5B; Table 2). This suggests that when the selective sweep of the A24R Leu18Phe variant occurred, these viruses also contained a single K3L CNV breakpoint. To determine whether the K3L gene copy number amplification or the A24R Leu18Phe mutation arose first in the virus population, we used Southern blot analysis to measure K3L CNV during experimental evolution of replicate A. We first detected K3L CNV at passage 4, and the proportion of viruses with the population harboring CNV seemed to remain steady through passage 10, as indicated by the consistent intensities of bands within each lane (Fig. 5C). Since the A24R Leu18Phe substitution did not emerge until passage 7, the earlier appearance of K3L CNV might have preemptively blocked PKR activation that would otherwise be induced by the A24R Leu18Phe variant, facilitating its rapid fixation. Indeed, in cells infected with the replicate A P10 virus, which contained both the A24R Leu18Phe substitution and K3L CNV, there was a marked reduction in PKR activation (Fig. 3E). This result suggests that K3L CNV can compensate for the activation of PKR induced by the A24R Leu18Phe variant.

To test whether rapid accumulation of the A24R Leu18Phe variant was influenced by the K3L copy number, we wanted to track increases of the A24R Leu18Phe variant in virus populations with and without K3L CNV. To do this, we infected cells with viruses containing the A24R Leu18Phe substitution alone (A24R^{Leu18Phe}), K3L CNV alone (CNV), or the combination of the two genetic changes (A24R^{Leu18Phe}+CNV), each starting at a ratio of 1:100 with the parental ΔE3L virus (see Materials and Methods for strain details). Compared to ΔE3L alone after two passages in HF cells, the A24R^{Leu18Phe}:ΔE3L virus population replicated ~10-fold better, CNV:ΔE3L ~1,000-fold better, and A24R^{Leu18Phe}:CNV:ΔE3L ~10,000-fold better (Fig. 5D). These data suggest that while either the A24R Leu18Phe variant or K3L CNV is sufficient for a fitness benefit in HF cells, the combination is additive and may therefore facilitate fixation of the A24R Leu18Phe variant. We next analyzed the abundance of the A24R Leu18Phe substitution in the virus populations to determine if it accumulated faster in the presence of K3L CNV. Starting at 1% of the population, the mutation reached 98.8% abundance after only two passages when multiple copies of K3L were present, compared to 45.2% abundance at passage 2 with a single copy of K3L (Fig. 5D). The increased accumulation of the variant in the presence of CNV is consistent with the added fitness benefit for viruses carrying both genetic changes. Thus, the A24R Leu18Phe substitution accumulated markedly faster in the presence of K3L CNV, which suggests that the reduction in PKR activation provided by increased K3L expression may have facilitated the rapid rise of the distant A24R mutation.

DISCUSSION

In this study, we found a new means of poxvirus adaptation to innate immune response pathways. Vaccinia virus has proven to be a useful model for experimental evolution and continues to reveal the genetic basis of various poxvirus adaptations (14, 20, 25). We charted the rise of multiple point mutations over the course of 10 serial infections, which is consistent with adaptive evolution through several independent mechanisms. Most notable among these was the rapid accumulation of a Leu18Phe substitution resulting from a mutation within A24R, which encodes a subunit of the viral RNA polymerase. Unlike in other evolved virus populations, the rise of the A24R

Leu18Phe variant appeared in the near absence of other point mutations (Fig. 5A), an observation reminiscent of clonal interference, where strongly adaptive point mutations on separate genomes can compete and transiently dominate within asexual populations (34, 35). However, a case of clonal interference seemed unlikely given the high rates of recombination in poxviruses (36–39) and the modest replicative advantage we measured in a recombinant strain containing only the A24R Leu18Phe mutation (Fig. 3A). We therefore considered the impact of K3L gene copy number amplification as a facilitating event for the rapid accumulation of the A24R Leu18Phe variant during this course of experimental evolution.

We previously found that genomes harboring multiple copies of K3L produced more protein product, which resulted in fitness gains for viruses under selective pressure to overcome the antiviral factor PKR (14). In two populations evolved in HeLa cells, we observed the additional emergence of a beneficial point mutation in K3L during serial infections, which we predict is more likely to occur in viruses with multiple copies of the gene (14). However, as we did not observe K3L mutations in every evolved population, adaptive mutations likely still arise at a low frequency. In this study, we observed a beneficial point mutation in a gene distant from the expanded K3L locus (>100,000 bp apart) that may benefit from the presence of virus genomes harboring multiple copies of K3L. We speculate that the Leu18Phe substitution encoded within the A24R variant alters virus transcription to promote virus replication and found that while the vRNAP variant does reduce total dsRNA, it also activates PKR (Fig. 3C and E). These results show that the A24R^{Leu18Phe} virus produces less dsRNA and suggest a complicated mechanism in which the polymerase variant alters RNA production in a way that activates PKR despite the reduction in dsRNA. Further analysis of viral transcripts may reveal changes to transcript length or other modifications to specific transcripts that modulate immune sensing.

Despite providing modest fitness gains and activating PKR, the A24R Leu18Phe mutation can rapidly sweep to fixation in virus populations in the presence of multiple copies of K3L that block PKR activation. Consistent with this idea, we observed rapid fixation of the A24R Leu18Phe variant following K3L copy number amplification (Fig. 5A and C) and a 2-fold increase in the accumulation of the A24R Leu18Phe variant in the presence of multiple K3L copies compared to that in populations with a single copy of K3L (Fig. 5D). These data suggest that adaptive copy number variation can facilitate the rapid accumulation of otherwise modestly beneficial mutations, such as the A24R Leu18Phe variant. In this way, recombination-based CNV may enhance the viability of an expanded set of beneficial mutations that otherwise suffer from trade-offs (e.g., activation of PKR) and might otherwise be unable to sweep through populations. Given that copy number amplification events are likely to be transient (14), this foothold may be temporary, as suggested by previous work describing the accumulation of beneficial point mutations causing the collapse of CNV (20). In the current study, CNV of the K3L locus persisted through passage 10 (Fig. 5C), but it might collapse to a single copy after further rounds of infection. In any case, copy number variation provides an opportunity for mutations to sweep rapidly through genes both undergoing CNV and distant from CNV, despite small initial fitness gains relative to those for other mutations in the population. Given long periods of evolutionary time, the presence of seemingly simple, beneficial point mutations in virus populations may belie a more volatile history of fixation involving the aid of adaptive yet transient gene copy number amplification events.

MATERIALS AND METHODS

Cells and viruses. Primary human fibroblast (HF) cells derived from human foreskin were a gift from Adam Geballe (Fred Hutchinson Cancer Research Center). HF, HeLa, and BHK cells were maintained in Dulbecco's modified Eagle's medium (DMEM; HyClone) supplemented with 10% fetal bovine serum (HyClone), 1% penicillin-streptomycin (GE Lifesciences), and 1% stable L-glutamine (GE Lifesciences). Rhesus fibroblasts (from *Macaca mulatta*; Coriell Institute for Medical Research) and African green monkey fibroblasts (from *Cercopithecus aethiops*; Coriell Institute for Medical Research) were maintained in minimum essential medium, alpha modification (MEM-alpha; HyClone), supplemented as described

above for DMEM. The Copenhagen strain of vaccinia virus (VC-2) and the E3L deletion virus (ΔE3L) (13) were generous gifts from Bertram Jacobs (Arizona State University).

Experimental evolution. For each infection during experimental evolution, 150-mm dishes were seeded with an aliquot from the same stock of HF cells (5×10^6 cells/dish). Triplicate dishes of cells were infected (multiplicity of infection [MOI] = 1.0 for P1 and 0.1 for subsequent passages) from a single stock of ΔE3L virus for 2 h in a minimal volume and then supplemented with medium. At 48 h, cells were collected, washed, pelleted, and resuspended in 1 ml of medium. Virus was released by one freeze-thaw cycle followed by sonication. Viral titers were determined by a 48-h plaque assay with BHK cells between passages. Following 10 passages, equal volumes of virus from every other passage were expanded in BHK cells for 48 h, with viral titers determined by a 48-h plaque assay with BHK cells, or for 72 h for replicate A passage 10 due to a small-plaque phenotype. An equal volume from the input parental ΔE3L virus stock was similarly expanded for comparison.

VACV whole-genome deep sequencing. Total viral genomic DNA was collected following a 24-h infection of BHK cells (MOI = 0.1) as previously described (41). Libraries were constructed using a Nextera XT DNA sample prep kit (Illumina, Inc.). Barcoded libraries were combined and sequenced using an Illumina MiSeq instrument at the High-Throughput Genomics Core (University of Utah). Reads were mapped to the VC-2 reference genome (accession number M35027.1; modified on http://poxvirus.org) (42) by using BWA mem (v0.7.10) (40) in default mode. PCR duplicates were removed, and the read depth was calculated using samtools (v0.1.18) (43). We utilized the Genome Analysis Toolkit (v3.2-2) (44) for base quality score recalibration, indel realignment, and variant calling across all samples (45, 46). We utilized Wham (v1.7.0-272-g078c-dirty) for structural variant calling (23). SNP and depth plots were generated in R (https://www.r-project.org/).

Recombinant virus generation. A sequence of 500 bp of homology surrounding the Leu18Phe-encoding mutation in A24R was amplified by PCR from replicate A passage 10 viral DNA by using primers A24R_1F (5′-CCTCCTCTCGAGCCCTCTCTGTTAGATGAGGATAGC) and A24R_1R (5′-CCTCCTACTAGTCAGTG AACGTGGCTAATGCG). A sequence of 500 bp of homology surrounding the Lys452Asn-encoding mutation in A24R was amplified by PCR from VC-2 viral DNA by using primers A24R_2F (5′-CCTCCTCTCGAG CGTTGGCACATGATGAATTAGAGAATTAC) and A24R_2R (5′-CCTCCTACTAGTGAGATGCGACTAGAGCATTTT CTATAGTG). The resulting PCR products were digested with XhoI and SpeI (New England BioLabs), gel purified, and cloned into pEQ1422 (a gift from A. Geballe, Fred Hutchinson Cancer Research Center) (20) cut with the same enzymes to generate pEQ1422-Leu18Phe and pEQ1422-A24R_2. The Lys452 Asn-encoding mutation was introduced into pEQ1422-A24R_2 by use of site-directed mutagenesis primers A24R_Lys452Asn_F (5′-GTTGGATTTTATCCGGATCAAGTAAATATTTCAAAGATGTTTTCTGTCA) and A24R_Lys452Asn_R (5′-TGACAGAAAACATCTTTGAAATATTTACTTGATCCGGATAAAATCCAAC) (pEQ1422-Lys452Asn).

BHK cells were infected with ΔE3L (MOI = 1.0) and then transfected at 1 h postinfection with pEQ1422-Leu18Phe or pEQ1422-Lys452Asn by use of FuGENE6 (Promega) according to the manufacturer's protocol. Infected cells were collected at 48 h postinfection, and viruses were released by one freeze-thaw cycle followed by sonication. Resulting viruses were selected using transient dominant selection (47). Briefly, viruses were plaque purified three times in the presence of 600 μg/ml hygromycin B (Sigma-Aldrich), followed by three rounds of plaque purification without hygromycin B. The presence of the Leu18Phe or Lys452Asn substitution and loss of the Hyg^r phenotype were confirmed by PCR followed by Sanger sequencing. Viruses were amplified and titers measured in BHK cells.

One-step growth curve. BHK or HF cells were infected with VACV ΔE3L or A24R^{Leu18Phe} (MOI = 5.0) in triplicate, and the virus was replaced with fresh medium after 2 h. Cells were harvested at 2, 6, 12, 24, 48, and 72 h postinfection, and viral titers were determined by a 72-h plaque assay with BHK cells.

Modeling. Amino acid alignment between the VACV A24R gene product and *S. cerevisiae* RBP2 was generated using Clustal Omega (v1.2.1). Corresponding A24R variant residues were then mapped onto the *S. cerevisiae* RNAP II crystal structure (PDB entry 1I50) (28) by use of Chimera software (48; http://www.cgl.ucsf.edu/chimera/).

dsRNA dot blotting. HF cells were left untreated (mock) or infected with the ΔE3L, A24R^{Leu18Phe}, or A24R^{Lys452Asn} virus (MOI = 10) for 13 h. Total RNA was collected from infected cells, and dilutions were spotted onto nylon membranes (GE Lifesciences). Membranes were allowed to dry before two rounds of UV cross-linking. Blots were blocked for 1 h in phosphate-buffered saline with Tween (PBST) plus 5% milk and then incubated with MαdsRNA J2 (1:1,000; Scicons) for 1 h followed by GαM-IgG-HRP (1:50,000; Millipore) for 1 h. Blots were then activated with WesternBright ECL reagent (Advansta) and exposed to autoradiography film (GeneMate), which was developed in a Mini-Med 90 film processor (AFP Imaging).

rRNA degradation assay. HF cells were pretreated with interferon alpha for 24 h. Cells were then left untreated (mock), transfected with poly(I-C), or infected with the ΔE3L, A24R^{Leu18Phe}, or A24R^{Lys452Asn} virus (MOI = 5.0) for 6 h. Total RNA was harvested, and a total of 225 ng/lane was run on an Agilent 2200 TapeStation instrument.

Immunoblot analysis. HF cells were left untreated (mock) or infected with the ΔE3L, A24R^{Leu18Phe}, A24R^{Lys452Asn}, or replicate A P10 virus (MOI = 5.0) for 6 h. Protein lysates were collected in RIPA lysis buffer, and total protein concentrations were quantified by Bradford assay by use of a Synergy HT plate reader (BioTek). Equivalent amounts of lysate were separated in a precast Mini-Protean TGX gel (Bio-Rad) and transferred to a polyvinylidene difluoride (PVDF) membrane (Immobilon). Blots were blocked for 30 min in PBST plus 5% milk, or for 10 min followed by three washes for phospho-specific antibodies. Blots were then incubated with the following primary antibodies overnight at 4°C: MαPKR B-10 (1:200; Santa Cruz), RαPhospho-PKR E120 (1:500; Abcam), RαeIF2α (1:1,000; Cell Signaling), and MαPhospho-eIF2α (1:250; Cell Signaling). Blots were probed with the appropriate secondary antibody for 1 h at room

temperature, i.e., GαM-IgG-HRP (1:50,000; Millipore) or GαR-IgG-HRP (1:50,000; Millipore), activated with WesternBright ECL reagent (Advansta), and exposed to autoradiography film (GeneMate), which was developed in a Mini-Med 90 film processor (AFP Imaging).

Plaque size analysis. BHK cells were infected with the ΔE3L or A24R^{Leu18Phe} virus (MOI = 0.1) for 48 h and then stained with crystal violet. Plates were imaged on a Gel Doc XR+ system (Bio-Rad), and plaque size was quantified using ImageJ v1.48 software (Rasband). Three independent wells, for a total of 350 plaques, were analyzed for each virus.

Southern blot analysis. Total viral DNA was collected as described above, and ~2 μg per sample was digested with BspEI (New England BioLabs). Digested DNAs were separated by agarose gel electrophoresis and transferred to nylon membranes (GE Lifesciences) by vacuum transfer, followed by UV cross-linking. The resulting blots were probed with PCR-amplified K3L by use of a DIG High-Prime DNA Labeling & Detection starter kit II (Roche) according to the manufacturer's protocol.

Mutation accumulation assay. The replicate A P10 virus was passaged an additional three times in HF cells until the A24R Leu18Phe variant was fixed (A24R^{Leu18Phe}+CNV virus). The K3L CNV-only virus contains the same CNV breakpoint but has no mutations in A24R (CNV virus). BHK cells were infected with viruses mixed at a ratio of 1:100, i.e., A24R^{Leu18Phe}:ΔE3L, CNV:ΔE3L, or A24R^{Leu18Phe}+CNV:ΔE3L, for two passages in BHK cells (MOI = 0.1). All passages were collected after 48 h as described above and titrated by a 72-h plaque assay with BHK cells.

Accession number(s). All deep-sequencing data are available at the Sequence Read Archive under accession number SRP073123.

ACKNOWLEDGMENTS

Experiments were designed by K.R.C. and N.C.E. Experiments were performed by K.R.C. Data were analyzed by K.R.C. and Z.N.K. Analysis tools were contributed by Z.N.K. and M.Y. The paper was written by K.R.C. and N.C.E.

We thank Adam Geballe (Fred Hutchinson Cancer Research Center) and Bertram Jacobs (University of Arizona) for reagents. We also thank Adam Geballe for valuable insights and critical assessment of the manuscript. We thank the High-Throughput Genomics Core (University of Utah) for technical assistance.

This study was supported by NIH grants R01GM114514 (N.C.E.), R01GM104390 (M.Y.), T32AI055434 (K.R.C.), and T32GM007464 (Z.N.K.). N.C.E. was supported by the Pew Biomedical Scholars program and the Mario R. Capecchi Endowed Chair in Genetics.

The experimental design, data collection and analysis, tools, and decision to submit for publication are our own, and independent of the National Institutes of Health.

REFERENCES

1. Duffy S, Shackelton LA, Holmes EC. 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet 9:267–276. https://doi.org/10.1038/nrg2323.
2. Gago S, Elena SF, Flores R, Sanjuan R. 2009. Extremely high mutation rate of a hammerhead viroid. Science 323:1308. https://doi.org/10.1126/science.1169202.
3. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. 2010. Viral mutation rates. J Virol 84:9733–9748. https://doi.org/10.1128/JVI.00694-10.
4. Lynch M. 2010. Evolution of the mutation rate. Trends Genet 26:345–352. https://doi.org/10.1016/j.tig.2010.05.003.
5. Challberg MD, Englund PT. 1979. Purification and properties of the deoxyribonucleic acid polymerase induced by vaccinia virus. J Biol Chem 254:7812–7819.
6. Qin L, Evans DH. 2014. Genome scale patterns of recombination between co-infecting vaccinia viruses. J Virol 88:5277–5286. https://doi.org/10.1128/JVI.00022-14.
7. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. 2010. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Mol Biol Evol 27:2038–2051. https://doi.org/10.1093/molbev/msq088.
8. Kerr PJ, Ghedin E, DePasse JV, Fitch A, Cattadori IM, Hudson PJ, Tscharke DC, Read AF, Holmes EC. 2012. Evolutionary history and attenuation of myxoma virus on two continents. PLoS Pathog 8:e1002950. https://doi.org/10.1371/journal.ppat.1002950.
9. Davies MV, Chang HW, Jacobs BL, Kaufman RJ. 1993. The E3L and K3L vaccinia virus gene products stimulate translation through inhibition of the double-stranded RNA-dependent protein kinase by different mechanisms. J Virol 67:1688–1692.
10. Chang HW, Watson JC, Jacobs BL. 1992. The E3L gene of vaccinia virus encodes an inhibitor of the interferon-induced, double-stranded RNA-dependent protein kinase. Proc Natl Acad Sci U S A 89:4825–4829. https://doi.org/10.1073/pnas.89.11.4825.
11. Davies MV, Elroy-Stein O, Jagus R, Moss B, Kaufman RJ. 1992. The vaccinia virus K3L gene product potentiates translation by inhibiting double-stranded-RNA-activated protein kinase and phosphorylation of the alpha subunit of eukaryotic initiation factor 2. J Virol 66:1943–1950.
12. Langland JO, Jacobs BL. 2002. The role of the PKR-inhibitory genes, E3L and K3L, in determining vaccinia virus host range. Virology 299:133–141. https://doi.org/10.1006/viro.2002.1479.
13. Beattie E, Denzler KL, Tartaglia J, Perkus ME, Paoletti E, Jacobs BL. 1995. Reversal of the interferon-sensitive phenotype of a vaccinia virus lacking E3L by expression of the reovirus S4 gene. J Virol 69:499–505.
14. Elde NC, Child SJ, Eickbush MT, Kitzman JO, Rogers KS, Shendure J, Geballe AP, Malik HS. 2012. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell 150:831–841. https://doi.org/10.1016/j.cell.2012.05.049.
15. van Boxel-Dezaire AHH, Rani MRS, Stark GR. 2006. Complex modulation of cell type-specific signaling in response to type I interferons. Immunity 25:361–372. https://doi.org/10.1016/j.immuni.2006.08.014.
16. Amegadzie BY, Holmes MH, Cole NB, Jones EV, Earl PL, Moss B. 1991. Identification, sequence, and expression of the gene encoding the second-largest subunit of the vaccinia virus DNA-dependent RNA polymerase. Virology 180:88–98. https://doi.org/10.1016/0042-6822(91)90012-Z.
17. Jones EV, Moss B. 1984. Mapping of the vaccinia virus DNA polymerase gene by marker rescue and cell-free translation of selected RNA. J Virol 49:72–77.

18. Traktman P, Sridhar P, Condit RC, Roberts BE. 1984. Transcriptional mapping of the DNA polymerase gene of vaccinia virus. J Virol 49: 125–131.
19. Betakova T, Wolffe EJ, Moss B. 1999. Regulation of vaccinia virus morphogenesis: phosphorylation of the A14L and A17L membrane proteins and C-terminal truncation of the A17L protein are dependent on the F10L kinase. J Virol 73:3534–3543.
20. Brennan G, Kitzman JO, Shendure J, Geballe AP. 2015. Experimental evolution identifies vaccinia virus mutations in A24R and A35R that antagonize the protein kinase R pathway and accompany collapse of an extragenic gene amplification. J Virol 89:9986–9997. https://doi.org/10.1128/JVI.01233-15.
21. Condit RC, Easterly R, Pacha RF, Fathi Z, Meis RJ. 1991. A vaccinia virus isatin-beta-thiosemicarbazone resistance mutation maps in the viral gene encoding the 132-kDa subunit of RNA polymerase. Virology 185: 857–861. https://doi.org/10.1016/0042-6822(91)90559-T.
22. Prins C, Cresawn SG, Condit RC. 2004. An isatin-beta-thiosemicarbazone-resistant vaccinia virus containing a mutation in the second largest subunit of the viral RNA polymerase is defective in transcription elongation. J Biol Chem 279:44858–44871. https://doi.org/10.1074/jbc.M408167200.
23. Kronenberg ZN, Osborne EJ, Cone KR, Kennedy BJ, Domyan ET, Shapiro MD, Elde NC, Yandell M. 2015. Wham: identifying structural variants of biological consequence. PLoS Comput Biol 11:e1004572. https://doi.org/10.1371/journal.pcbi.1004572.
24. Slabaugh MB, Roseman NA, Mathews CK. 1989. Amplification of the ribonucleotide reductase small subunit gene: analysis of novel joints and the mechanism of gene duplication in vaccinia virus. Nucleic Acids Res 17:7073–7088. https://doi.org/10.1093/nar/17.17.7073.
25. Brennan G, Kitzman JO, Rothenburg S, Shendure J, Geballe AP. 2014. Adaptive gene amplification as an intermediate step in the expansion of virus host range. PLoS Pathog 10:e1004002. https://doi.org/10.1371/journal.ppat.1004002.
26. Erlandson KJ, Cotter CA, Charity JC, Martens C, Fischer ER, Ricklefs SM, Porcella SF, Moss B. 2014. Duplication of the A17L locus of vaccinia virus provides an alternate route to rifampin resistance. J Virol 88: 11576–11585. https://doi.org/10.1128/JVI.00618-14.
27. Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, Thompson NE, Burgess RR, Edwards AM, David PR, Kornberg RD. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. Science 288:640–649. https://doi.org/10.1126/science.288.5466.640.
28. Cramer P, Bushnell DA, Kornberg RD. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science 292:1863–1876. https://doi.org/10.1126/science.1059493.
29. Cohrs RJ, Condit RC, Pacha RF, Thompson CL, Sharma OK. 1989. Modulation of ppp(A2′p)nA-dependent RNase by a temperature-sensitive mutant of vaccinia virus. J Virol 63:948–951.
30. Bayliss CD, Condit RC. 1993. Temperature-sensitive mutants in the vaccinia virus A18R gene increase double-stranded RNA synthesis as a result of aberrant viral transcription. Virology 194:254–262. https://doi.org/10.1006/viro.1993.1256.
31. Sanderson CM, Frischknecht F, Way M, Hollinshead M, Smith GL. 1998. Roles of vaccinia virus EEV-specific proteins in intracellular actin tail formation and low pH-induced cell-cell fusion. J Gen Virol 79:1415–1425. https://doi.org/10.1099/0022-1317-79-6-1415.
32. Röttger S, Frischknecht F, Reckmann I, Smith GL, Way M. 1999. Interactions between vaccinia virus IEV membrane proteins and their roles in IEV assembly and actin tail formation. J Virol 73:2863–2875.
33. Wolffe EJ, Weisberg AS, Moss B. 1998. Role for the vaccinia virus A36R outer envelope protein in the formation of virus-tipped actin-containing microvilli and cell-to-cell virus spread. Virology 244:20–26. https://doi.org/10.1006/viro.1998.9103.
34. Fisher RA. 1930. The genetical theory of natural selection. Clarendon Press, Oxford, United Kingdom.
35. Muller HJ. 1932. Some genetic aspects of sex. Am Nat 66:118–138. https://doi.org/10.1086/280418.
36. Ball LA. 1987. High-frequency homologous recombination in vaccinia virus DNA. J Virol 61:1788–1795.
37. Evans DH, Stuart D, McFadden G. 1988. High levels of genetic recombination among cotransfected plasmid DNAs in poxvirus-infected mammalian cells. J Virol 62:367–375.
38. Spyropoulos DD, Roberts BE, Panicali DL, Cohen LK. 1988. Delineation of the viral products of recombination in vaccinia virus-infected cells. J Virol 62:1046–1054.
39. Merchlinsky M. 1989. Intramolecular homologous recombination in cells infected with temperature-sensitive mutants of vaccinia virus. J Virol 63:2030–2035.
40. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
41. Esposito J, Condit R, Obijeski J. 1981. The preparation of orthopoxvirus DNA. J Virol Methods 2:175–179. https://doi.org/10.1016/0166-0934(81)90036-7.
42. Goebel SJ, Johnson GP, Perkus ME, Davis SW, Winslow JP, Paoletti E. 1990. The complete DNA sequence of vaccinia virus. Virology 179: 247–266, 517–563. https://doi.org/10.1016/0042-6822(90)90294-2.
43. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509.
44. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110.
45. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498. https://doi.org/10.1038/ng.806.
46. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 11:11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43.
47. Falkner FG, Moss B. 1990. Transient dominant selection of recombinant vaccinia viruses. J Virol 64:3108–3111.
48. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20084.

CHAPTER 4

LONG READ NANOPORE SEQUENCING REVEALS CONCERTED EVOLUTION OF A BENEFICIAL POINT MUTATION IN VACCINIA VIRUS

Kelsey R. Cone*[1], Thomas A. Sasani*[1,2], Aaron R. Quinlan[1,2], and Nels C. Elde[1]

Abstract

Gene copy number variation promotes rapid adaptation of poxviruses. These gene amplification events can also involve the emergence of beneficial single nucleotide variants that accumulate in dynamic gene arrays. Tracking the dynamics of variant fixation in heterogeneous virus populations is a difficult task, especially when relying solely on short read sequencing technologies. We developed methods using Oxford Nanopore long read sequencing to perform high resolution analyses of single nucleotide variation within a poxvirus host range gene undergoing copy number amplification during courses of experimental evolution. These methods allowed us to phase a beneficial point mutation generating a His47Arg variant within expanding K3L gene arrays in individual viral genomes sampled from the population. Our results reveal concerted

[1] Department of Human Genetics, University of Utah, Salt Lake City, Utah, U.S.A.
[2] USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, Utah, U.S.A.
* Equal contributors

evolution of the K3L^{His47Arg} variant, in which virus genomes harboring multiple K3L copies rapidly homogenized the beneficial mutation. These intriguing observations of poxvirus adaptation through concerted evolution reveal a new facet of virus evolution reminiscent of other organisms containing large repetitive genomic regions.

Introduction

Viruses can undergo rapid adaptation, given large population sizes and short replication cycles facilitating the accumulation of beneficial variants. For RNA viruses, high point mutation rates generate diverse populations from which advantageous variants quickly emerge under selection (Duffy et al., 2008; Holland et al., 1982; Sanjuán et al., 2010). Studies tracking RNA viruses revealed the simultaneous accumulation of mutations in a population, described as quasi-species, that reflect the success of highly heterogeneous virus populations (reviewed in Andino and Domingo, 2015; Domingo et al., 2012). Many DNA viruses contain larger genomes and lower point mutation rates (Gago et al., 2009; Sanjuán et al., 2010), which makes tracking variants at a population level more difficult. As such, less is known about the genetic diversity maintained in populations of DNA viruses.

The DNA virus family *Poxviridae* presents a useful system for examining population dynamics of variant accumulation because of high recombination rates that may allow for advantageous mutations to emerge simultaneously within individual genomes (Ball, 1987; Evans et al., 1988; Merchlinsky, 1989; Spyropoulos et al., 1988). Poxvirus DNA polymerase genes encode both replicase and recombinase activities, reflecting a tight coupling of these essential functions for virus replication (Colinas et al.,

1990; Hamilton and Evans, 2005; Willer et al., 1999). Abundant recombination may also reduce the impact of clonal interference between genomes (Kim, 2005; McDonald et al., 2016), and could facilitate the spread of rare variants between co-infecting viruses (Xiao et al., 2016). Poxvirus recombination also influences structural variation within genomes, which can result in rapid gene copy number amplification events (Brennan et al., 2014; Cone et al., 2017; Elde et al., 2012). Copy number variation (CNV) is also proposed to provide a genetic foothold in otherwise unfavorable conditions to allow greater sampling of potentially beneficial mutations (Brennan et al., 2014) distributed both within and outside of genomic regions undergoing CNV (Cone et al., 2017). Moreover, the presence of gene copy arrays predicts the increased probability of a beneficial point mutation arising in the expanded locus. Consistent with this idea, we discovered an advantageous single nucleotide variant (SNV) within amplified K3L gene arrays in populations of vaccinia virus during courses of serial infection (Elde et al., 2012). Although the beneficial point mutation, resulting in a His47Arg amino acid change ($K3L^{His47Arg}$), emerged in multiple replicate populations, the fate of the variant within virus genomes harboring copy number amplification of K3L is unclear.

The emergence of a single nucleotide variant within a gene undergoing copy number variation in an evolving poxvirus population provides a unique system to explore variant dynamics in the context of recombination and gene amplification events. It has long been postulated that gene duplications are an important source for creating genetic novelty (Nei, 1969; Ohno, 1970; Stephens, 1951), and studies in eukaryotic and bacterial systems have revealed some of the ways gene duplications promote evolution (reviewed in Magadum et al., 2013). Whether similar patterns exist in viral genome evolution is

unknown. In this study, analyses of variant dynamics during gene copy number amplifications in poxvirus populations reveal new insights into adaptive mechanisms driving poxvirus genome evolution. We developed methods to apply Oxford Nanopore long read sequencing to discover the precise location of the K3L^{His47Arg} single nucleotide variant within K3L gene arrays. The ability to phase variants within individual virus genomes containing multiple K3L gene copies revealed patterns of variant accumulation similar to those found in other organisms such as bacteria and yeast. These observations imply that certain aspects of genome evolution are shared between viruses and their hosts.

Results

Accumulation of a single nucleotide variant within gene copy number expansions

To explore variant dynamics in a virus population following gene amplification, we studied a vaccinia virus (VACV) population in which gene copy number variation preceded the appearance of a beneficial point mutation in that gene. Adaptive CNV of the VACV K3L gene was detected after four serial passages under selection in HeLa cells, and overall gene copy number appeared fairly constant by passage ten (P10) (Elde et al., 2012). The SNV generating the K3L^{His47Arg} variant was first detected at P5, and reached a population allele frequency of about 12% by P10, suggesting that the variant confers a small fitness benefit (Elde et al., 2012). Consistent with this idea, the K3L^{His47Arg} variant was previously shown to more potently inhibit human PKR (Kawagishi-Kobayashi et al., 1997). To determine whether the K3L^{His47Arg} variant would continue to increase in

abundance under selection, we performed an additional ten passages in triplicate in HeLa cells starting with the P10 virus population. All three replicate populations showed additional fitness gains through P14, at which point viral titers reached a plateau that continued through P20 (Figure 4.1). These results suggest all virus populations reached a maximum replication ability under these selective conditions.

We analyzed viral genomes from each P20 population by deep sequencing on an Illumina MiSeq instrument to assess genetic changes underlying adaptation, and found that the K3L^{His47Arg} variant had increased in frequency to near fixation in all three replicate populations (frequency = 0.86-0.91, Figure 4.1). Further analysis of the replicate A population revealed that the K3L^{His47Arg} variant reached an intermediate frequency at P15 (frequency = 0.58, Figure 4.1), consistent with a steady accumulation of the K3L^{His47Arg} variant over time. We next assessed the same replicate A virus populations for the presence of K3L CNV to determine whether there were population-level changes in K3L copy number. Southern blot analyses of passages 10, 15, and 20 revealed a similar proportion of viruses harboring multiple copies of K3L throughout these passages (Figure 4.1). Thus, we did not observe any changes to overall K3L CNV, consistent with a constant selective pressure maintaining multiple K3L copies within the population. These data show that the K3L^{His47Arg} variant continued to increase in frequency from P10 to P20, while overall gene copy number distribution remained constant.

Identification of single-copy and multicopy K3L sequencing reads

We utilized an in-depth analysis of viral genomes containing different K3L copy number to determine the dynamics by which the K3L^{His47Arg} variant accumulated within

expanded gene arrays. The two structural variant breakpoints identified in the P10 population generating tandem K3L arrays (Elde et al., 2012) each create a repetitive unit over 500bp in length (Figure 4.2). Thus, our initial 150bp paired-end Illumina MiSeq sequencing reads were of insufficient length to definitively determine whether the $K3L^{His47Arg}$ variant is present in multicopy genomes (Figure 4.2). To identify single nucleotide variants within genomes containing multiple K3L copies, we utilized long reads generated on the ONT MinION platform. The MinION sequencing platform is capable of reliably generating 6-8kbp reads, and is reported to produce reads in excess of 150kbp (Ip et al., 2015; Jain et al., 2016). Thus, a single read capable of spanning multiple K3L copies permits variant calling in a multicopy genome, which has not previously been feasible for these large DNA viruses using current Illumina sequencing technologies (Figure 4.2).

In our studies, long sequencing reads allowed us to phase the $K3L^{His47Arg}$ variant to single viral genomes in the context of K3L CNV. We performed multiple independent MinION sequencing reactions with viral genomes from each of the replicate A P10, P15, and P20 populations. For each population, we generated ~40,000 ONT reads that mapped to the VC-2 reference genome (Table 4.1), with a mean read length of about 4kbp. To identify reads containing K3L, we selected all reads that aligned to the repetitive unit containing K3L (Figure 4.2) at least once. Variant calling using this ONT read set showed that population-level $K3L^{His47Arg}$ variant frequencies were nearly identical to those estimated using prior Illumina MiSeq data (Figure 4.1). Total coverage at the $K3L^{His47Arg}$ variant site was approximately equal in both data sets (Table 4.1). Thus, despite an error rate of approximately 8% (Jain et al., 2016), ONT long read sequencing

accurately and reliably identified a single nucleotide variant within VACV genomes.

We next categorized ONT reads containing K3L into single-copy or multicopy reads. Reads that unambiguously mapped to the K3L locus a single time were classified as single-copy. Since the VC-2 reference sequence itself contains a single K3L gene, if more than one distinct portion of a read aligned to the full length of the K3L repetitive unit, it was instead classified as multicopy (Figure 4.2). We then further selected reads that mapped to unique sequence in the VC-2 reference genome at least 75bp upstream and downstream of the repetitive unit, and thus encompass the total number of K3L copies from a given viral genome. These methods resulted in a set of reads for replicate A populations (ONT-K3L CNV read set) that represents a snapshot of K3L copy number in viral genomes at each passage analyzed.

We utilized the ONT-K3L CNV read set from P10, P15, and P20 to gain a detailed understanding of K3L copy number within the different virus populations. The majority of reads contain between one and five copies of K3L, and this distribution does not change between the P10, P15, and P20 populations (Figure 4.2). These data are consistent with our Southern blot analyses (Figure 4.1), and suggest that a dynamic recombination process can generate stable copy number increases at the population level. Given our average read length of ~4kbp, we may be limited in our discovery of viral genomes with greater than 10 K3L copies. Specific long read DNA library preparations could allow for increased read length in the future. However, we did identify some reads containing high K3L copy number within the ONT-K3L CNV read set, with a maximum of 21 K3L gene copies in a single read (Figure 4.2). The largest of these tandem arrays correspond to significant increases in genome size, which could incur a fitness cost to

maintaining an extremely large genome. However, genomic additions of similar size have been shown to have little impact on replication efficiency (Smith and Moss, 1983). Thus, while an excess of sequence is tolerated, the bias towards low copy number is consistent with an equilibrium between the benefits of extra gene copies and the cost of a large genome size.

Concerted evolution of a point mutation within viral genomes

We next analyzed the ONT-K3L CNV read set for the presence of the K3L^{His47Arg} variant within genomes of varying K3L copy number across P10, P15, and P20 populations. Our analysis demonstrates that the variant frequency within reads containing up to five K3L copies closely resembles the population frequency at each passage (Figure 4.3). Although the vast majority of reads (~90%) contain between one and five copies of K3L, we observed similar distributions of the SNV within reads containing greater than five K3L copies (data not shown). Thus, our analysis of the ONT-K3L CNV read set suggests that there is no overall bias of the K3L^{His47Arg} variant towards lower or higher copy number genomes. One outlier is within the P10 population, where we observed that single-copy reads contained the K3L^{His47Arg} variant at a greater frequency than the population. This suggests that the K3L^{His47Arg} variant may have arisen in a single-copy virus, and then spread across multicopy viruses via recombination.

To determine exactly how a SNV spreads through tandem gene duplications within viral genomes, we identified the precise order of wildtype (WT) and variant alleles of K3L in multicopy reads. Utilizing the ONT-K3L CNV read set from P10, P15, and P20 populations, we identified the presence or absence of the K3L^{His47Arg} SNV in each

K3L copy of a multicopy read. As expected based on the population-level variant frequencies, we observed that most reads contain all WT K3L copies at P10, a split between WT and variant K3L arrays at P15, and all variant K3L copies at P20, independent of total copy number in the read (Figure 4.3). In all populations, there is a bias towards multicopy genomes harboring homogenous K3L arrays, in which genomes contain either all WT or all variant copies of K3L, although we also observe some multicopy genomes with mixed combinations of K3L alleles (Figure 4.3). If the accumulation of the $K3L^{His47Arg}$ variant were random, one would expect most multicopy genomes to contain a mixed set of WT and variant alleles, as there are more possible combinations of mixed alleles than homogenous alleles. Instead, we see a striking "all-or-none" scenario in all populations. Even at P15, when the variant is present at an intermediate frequency, multicopy genomes with entirely WT or variant alleles predominate. This pattern is reminiscent of concerted evolution, a molecular process leading to the homogenization of DNA sequences belonging to a repetitive gene family (Dover et al., 1982). Concerted evolution has been observed in many organisms (reviewed in Elder and Turner, 1995; Liao, 1999), but has not previously been observed in viruses. We suspect that the unique recombination abilities of poxviruses drive rapid homogenization of these gene arrays, allowing concerted evolution to proceed. Vaccinia virus could thus provide a tractable model in which to test the mechanisms driving this complex evolutionary process.

Recombination affects concerted evolution to drive adaptive outcomes

Concerted evolution in these virus populations is likely dependent upon recombination dynamics, which can affect adaptive outcomes for the population. The rapid homogenization of the K3L^{His47Arg} variant within gene arrays through recombination could occur either between replicating genomes from a single infection (intragenomic), or between co-infecting viruses (intergenomic). To test whether the pattern of concerted evolution is affected by changes in intergenomic recombination rates, we altered the multiplicity of infection (MOI) during passaging. We repeated P10 to P15 using a 4-log difference in MOI (1.0 - 0.001), to determine whether increasing or virtually eliminating the chance of intergenomic recombination would affect the way the K3L^{His47Arg} variant accumulated within the virus population. All populations show similar gains in fitness following passaging, independent of MOI (Figure 4.4). This result suggests that adaptation to the selective pressure conditions is robust to changes in intergenomic recombination rates. Examination of P15 viral genomes from each population using ONT sequencing will allow us to test whether accumulation of the K3L^{His47Arg} variant tracks with increases in fitness. These analyses should reveal whether concerted evolution is affected by or robust to changes in intergenomic recombination, and provide insight into the contribution of intragenomic recombination to the process.

The rapid homogenization of the K3L^{His47Arg} variant within gene arrays in response to selection in HeLa cells could provide immediate fitness benefits, as well as alter future adaptive outcomes. Genetic changes beneficial in one environment can positively or negatively affect the way a virus population adapts to different selective pressure conditions. For example, we previously observed a collapse of K3L copy

number in the P10 population under relaxed selective pressure, and most individual clones contained a single allele of K3L, whether WT or K3L^{His47Arg} (Elde et al., 2012, Figure 4.4). These data suggest that low K3L copy number is favored in these conditions, and multicopy virus clones are biased against a mixed set of K3L alleles. This implies that at P10, concerted evolution may already have played a role in the potential viral clones that can emerge from a diverse starting population. To test whether the presence of the K3L^{His47Arg} variant at higher frequencies affects the population dynamics when selective pressure is altered, we performed the same experiment with the P15 and P20 populations. Plaque purification of individual viral clones in a permissive cell line to relax selective pressure revealed a similar bias towards single or low-copy K3L genomes with a single K3L allele, but an increase in the proportion of clones harboring the K3L^{His47Arg} variant (Figure 4.4). The abundance of K3L^{His47Arg} clones following plaque purification closely matches the population-level K3L^{His47Arg} variant frequency for each passage (Figures 4.1, 4.4). This suggests that under relaxed selection, while low K3L copy number is favored, there is no bias towards the WT or variant allele. Thus, the effect of concerted evolution is to change the virus population structure, which in turn alters the adaptive potential in new environments.

Discussion

In this study, we tracked the accumulation dynamics of a point mutation in the context of recombination-mediated gene copy number variation during poxvirus evolution. Since poxviruses have relatively low point mutation rates, gene amplification events are proposed to provide a genetic foothold until further adaptations can occur

(Brennan et al., 2014). One such example is a VACV population in which an adaptive single nucleotide variant in the K3L gene appeared following K3L CNV (Elde et al., 2012). We utilized this population (P10) to study the interplay between different types of genetic adaptations during additional rounds of selection. Analyses of evolving populations using ONT long read sequencing methods revealed new information into SNV dynamics in the context of complex genome structures and recombination effects. With ONT reads, we were able to precisely determine K3L copy number distribution within three viral populations isolated during the course of experimental evolution. Interestingly, while additional fitness gains were observed from P10 to P20 (Figure 4.1), the distribution of K3L copy number in the viral populations remained relatively constant (Figures 4.1, 4.2). In contrast, the K3L^{His47Arg} variant increased in frequency to near fixation by P20, confirmed by near-identical Illumina and ONT-derived variant frequency estimates (Figure 4.1). Together, these results suggest that while there may be an upper limit to K3L array size that was already established by P10, additional adaptations in the form of point mutations can support continued fitness increases in the population.

Our analyses of precise gene copy number and SNV accumulation dynamics were only made possible through the use of the Oxford Nanopore system to generate long reads capable of spanning complex gene arrays. Other next-generation sequencing technologies provide greater depth of coverage, yet despite the increased error rates associated with ONT reads, we obtained similar estimates of K3L^{His47Arg} variant frequency in three vaccinia populations using either the Illumina MiSeq or ONT MinION platform. These results highlight the utility of ONT sequencing methods for detecting

both SNVs and structural variants, which is critical for our analyses of poxvirus populations in which recombination as well as point mutations drive rapid adaptation. ONT long reads provided a level of detail not possible with short read sequencing technologies, and the methods used in this study provide a framework to explore mutation and copy number dynamics in other viruses as well. In particular, analyses of large DNA virus populations have been limited compared to RNA viruses, in part due to differences in genome size. Our work demonstrates the ability of Oxford Nanopore sequencing to generate long, high-quality reads with enough yield to investigate DNA virus genome evolution in more detail.

In this study, the accumulation of the K3L^{His47Arg} variant in a vaccinia virus population with stable overall K3L copy number provided a unique data set in which to investigate the exact dynamics by which a SNV accumulates in the context of CNV. ONT sequencing reads enabled the discovery that selection for the K3L^{His47Arg} variant is independent of gene copy number, since the variant frequency within reads containing different K3L copy number closely matches the population-level variant frequency at each passage (Figure 4.3). Additionally, our analysis of the pattern of WT and K3L^{His47Arg} variant alleles within multicopy reads revealed a bias towards homogenous K3L arrays, consistent with a pattern of concerted evolution (Figure 4.3). At very low (P10) or high (P20) variant frequencies, one might expect to observe this “all-or-none” pattern. However, even at an intermediate variant frequency in the P15 population, multicopy genomes favored either all WT or all K3L^{His47Arg} copies. Thus, homogenous arrays predominate, despite the possibility of recombination between viral genomes, evidenced by a variety of allele combinations observed at low abundance in all populations.

Our studies revealed the first instance of concerted evolution in a virus population, and we propose VACV as a useful model to delve into the mechanisms driving this complex evolutionary process. We demonstrated the utility of the model by testing how changes to different experimental parameters, namely intergenomic recombination and changes in selective pressure, affect the adaptive outcomes for evolving virus populations (Figure 4.4). Altering the rate of intergenomic recombination did not affect the ability of virus populations to adapt to the cell line (Figure 4.4), but further work is needed to determine whether the $K3L^{His47Arg}$ variant accumulates as fitness increases. Varying the multiplicity of infection while maintaining all other variables should allow us to determine whether concerted evolution of the $K3L^{His47Arg}$ variant is dependent on the intergenomic recombination rate. We also showed that concerted evolution changes the population structure such that the adaptive potential in other environments is affected. Following relaxed selection, we observed that while all individual viral clones contained lower K3L copy number, the proportion of clones containing the adaptive $K3L^{His47Arg}$ variant correlated with the starting population frequency. Thus, the effect of gene array homogenization is to increase the chance of maintaining a variant allele in a different environment. These changes to the population structure can impact viral success when entering a new environment, such as during host-switching events, including zoonoses.

In systems ranging from bacteria to mammals, concerted evolution is observed in gene families (Liao, 1999), which is not a common feature among viruses. However, there are a few examples of gene families in DNA viruses, including poxviruses (McLysaght et al., 2003), herpesviruses (Searles et al., 1999), and giant viruses like

Mimivirus and Marseillevirus (Colson and Raoult, 2010). Our results demonstrating concerted evolution in VACV thus provide a model system to understand how this process could affect adaptation in these DNA viruses. This is particularly relevant for avipoxviruses, as up to 30% of the genome is composed of gene families (Afonso et al., 2000). Thus, in addition to understanding variant dynamics during adaptive gene amplifications, studies of concerted evolution specific to poxvirus gene families could reveal new aspects of poxvirus evolution. Mechanisms elucidated using VACV could likely be applied widely, since repetitive sequences evolving in a concerted manner is a general phenomenon within diverse systems, and VACV provides a tractable model system in which to study the complex mechanisms driving concerted evolution.

Materials and methods

Cells and viruses

HeLa and BHK cells were maintained in Dulbecco's modified Eagle's medium (DMEM; HyClone) supplemented with 10% fetal bovine serum (HyClone), 1% penicillin-streptomycin (GE Lifesciences), and 1% stable L-glutamine (GE Lifesciences).

The P10 strain of vaccinia virus was previously established following 10 serial passages of the ΔE3L virus (Beattie et al., 1995) in HeLa cells (replicate C passage 10 in Elde et al., 2012). P10 virus was expanded in BHK cells, and titer determined by 48-hour plaque assay in BHK cells performed in triplicate.

Experimental evolution

Passages were performed as in (Elde et al., 2012), starting with the P10 virus population. For each passage, triplicate 150mm dishes were seeded with an aliquot from the same stock of HeLa cells ($5x10^6$ cells/dish) and infected (MOI = 0.001 – 1.0 as indicated) for 48 hours. Cells were then collected, washed, pelleted, and resuspended in 1mL of media. Virus was released by one freeze/thaw cycle followed by sonication. Viral titers were determined between each passage by 48-hour plaque assay in BHK cells performed in triplicate. Following passage 20, titers were reassessed simultaneously by 48-hour plaque assay in BHK cells performed in triplicate.

Deep sequencing of VACV genomes

Illumina. Total viral genomic DNA was collected as previously described (Esposito et al., 1981) from infected BHK cells (MOI = 0.1) 24 hours post-infection. Libraries were constructed using the Nextera XT DNA sample prep kit (Illumina, Inc.). Barcoded libraries were pooled and sequenced on an Illumina MiSeq instrument at the High-Throughput Genomics Core (University of Utah). Reads were mapped to the Copenhagen reference strain of vaccinia virus (VC-2; accession M35027.1; modified on poxvirus.org) (Goebel et al., 1990) using BWA-MEM (v0.7.10) (Li and Durbin, 2009) in default mode. PCR duplicates were removed using samtools (v0.1.18) (Li 2011). Base quality score recalibration, indel realignment, and variant calling across all samples was performed using Genome Analysis Toolkit (v3.2-2) (DePristo et al., 2011; McKenna et al., 2010; Van der Auwera et al., 2013).

Oxford Nanopore Technologies. Viral particles were isolated from infected BHK

cells (MOI = 1.0) 24 hours post-infection, and viral cores were purified by ultracentrifugation through a 36% sucrose cushion at 18,000 rpm for 80 minutes. Total viral genomic DNA was extracted from purified viral cores as above (Esposito et al., 1981). All sequencing libraries were constructed using an Oxford Nanopore SQK-NSK007 sample prep kit, and sequenced on a MinION Mk1B device using R9 chemistry MinION flow cells (Oxford Nanopore Technologies Ltd.). Sequencing reactions were run for 48 hours, and sequenced reads were base-called using the Metrichor cloud-based software suite. FASTA and FASTQ sequences were extracted from raw FAST5 files using *poretools* (Loman and Quinlan, 2014). Pooled FASTQ reads for each sample were aligned to the VC-2 reference genome as above with BWA-MEM, using the default settings provided by the *ont2d* flag (Li and Durbin, 2009).

Custom Python scripts (github.com/tomsasani/concerted-evolution-vacv) were utilized to calculate both K3L copy number and $K3L^{His47Arg}$ variant frequency within aligned viral reads. Briefly, reads that aligned at least once to the full length of the K3L repetitive unit (R.U.) were extracted, and further selected for reads that align at least 75bp upstream and downstream of the unit. K3L copy number was calculated within each read by summing the total number of alignments to the R.U., including both primary and supplementary alignments. Presence of a WT or $K3L^{His47Arg}$ allele was identified for each K3L gene copy within these complete K3L reads, excluding reads that lacked an alignment to the reference sequence at the SNV location. All Python code used to determine copy number and identify the $K3L^{His47Arg}$ SNV is available in the above GitHub repository, in addition to all scripts used for plotting/visualizing the data.

Southern blot analysis

Viral DNA from purified viral cores was digested with EcoRV (New England Biolabs), and separated by agarose gel electrophoresis. DNA was transferred to nylon membranes (GE Lifesciences) using a vacuum transfer, followed by UV-crosslinking. Blots were probed with PCR-amplified K3L using the DIG High-Prime DNA Labeling & Detection Starter Kit II (Roche) according to the manufacturer's protocol.

Plaque purification

BHK cells were infected for 48 hours with dilutions of replicate A P15 and P20 viruses, and overlayed with 0.4% agarose. Single plaques were harvested and transferred to new BHK dishes, and resulting wells harvested after 48 hours. Virus was released by one freeze/thaw cycle followed by sonication. The process was repeated three additional times for a total of four plaque purifications. Viral DNA was extracted from 9 single clones from each population as above (Esposito et al., 1981), and assessed for K3L CNV and the K3L^{His47Arg} SNV by PCR and Sanger sequencing.

References

Afonso, C.L., Tulman, E.R., Lu, Z., Zsak, L., Kutish, G.F., and Rock, D.L. (2000). The genome of fowlpox virus. J. Virol. *74*, 3815-3831.

Andino, R., and Domingo, E. (2015). Viral quasispecies. Virology *479-480*, 46–51.

Ball, L.A. (1987). High-frequency homologous recombination in vaccinia virus DNA. J. Virol. *61*, 1788–1795.

Beattie, E., Denzler, K.L., Tartaglia, J., Perkus, M.E., Paoletti, E., and Jacobs, B.L. (1995). Reversal of the interferon-sensitive phenotype of a vaccinia virus lacking E3L by expression of the reovirus S4 gene. J. Virol. *69*, 499–505.

Brennan, G., Kitzman, J.O., Rothenburg, S., Shendure, J., and Geballe, A.P. (2014). Adaptive gene amplification as an intermediate step in the expansion of virus host range. PLoS Pathog. *10*, e1004002.

Colinas, R.J., Condit, R.C., and Paoletti, E. (1990). Extrachromosomal recombination in vaccinia-infected cells requires a functional DNA polymerase participating at a level other than DNA replication. Virus Res. *18*, 49–70.

Colson, P., and Raoult, D. (2010). Gene repertoire of amoeba-associated giant viruses. Intervirology *53*, 330-343.

Cone, K.R., Kronenberg, Z.N., Yandell, M., and Elde, N.C. (2017). Emergence of a viral RNA polymerase variant during gene copy number amplification promotes rapid evolution of vaccinia virus. J. Virol. *91*, e01428-16.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., and Daly, M.J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

Domingo, E., Sheldon, J., and Perales, C. (2012). Viral quasispecies evolution. Microbiol. Mol. Biol. Rev. *76*, 159–216.

Dover, G. A., Brown, S., Coen, E., Dallas, J., Strachan, T., and Trick, M. (1982). The dynamics of genome evolution and species differentiation. In Genome Evolution, G.A. Dover and R.B. Flavell, eds. (Academic Press), p. 343.

Duffy, S., Shackelton, L.A., and Holmes, E.C. (2008). Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. *9*, 267–276.

Elde, N.C., Child, S.J., Eickbush, M.T., Kitzman, J.O., Rogers, K.S., Shendure, J., Geballe, A.P., and Malik, H.S. (2012). Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell *150*, 831–841.

Elder, J.F. Jr., and Turner, B.J. (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. Q. Rev. Biol. *70*, 297–320.

Esposito, J., Condit, R., and Obijeski, J. (1981). The preparation of Orthopoxvirus DNA. J. Virol. Methods *2*, 175–179.

Evans, D.H., Stuart, D., and McFadden, G. (1988). High levels of genetic recombination among cotransfected plasmid DNAs in poxvirus-infected mammalian cells. J. Virol. *62*, 367–375.

Gago, S., Elena, S.F., Flores, R., and Sanjuán, R. (2009). Extremely high mutation rate of

a hammerhead viroid. Science *323*, 1308.

Goebel, S.J., Johnson, G.P., Perkus, M.E., Davis, S.W., Winslow, J.P., and Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. Virology *179*, 247–66, 517–63.

Hamilton, M.D., and Evans, D.H. (2005). Enzymatic processing of replication and recombination intermediates by the vaccinia virus DNA polymerase. Nucleic Acids Res. *33*, 2259–2268.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and VandePol, S. (1982). Rapid evolution of RNA genomes. Science *215*, 1577–1585.

Ip, C.L., Loose, M., Tyson, J.R., de Cesare, M., Brown, B.L., Jain, M., Leggett, R.M., Eccles, D.A., Zalunin, V., Urban, J.M., Piazza, P., Bowden, R.J., Paten, B., Mwaigwisya, S., Batty, E.M., Simpson, J.T., Snutch, T.P., Birney, E., Buck, D., Goodwin, S., Jansen, H.J., O'Grady, J., Olsen, H.E., and MinION Analysis and Reference Consortium. (2015). MinION analysis and reference consortium: phase 1 data release and analysis. F1000Res. *4*, 1075.

Jain, M., Olsen, H.E., Paten, B., and Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. *17*, 239.

Kawagishi-Kobayashi, M., Silverman, J.B., Ung, T.L., and Dever, T.E. (1997). Regulation of the protein kinase PKR by the vaccinia virus pseudosubstrate inhibitor K3L is dependent on residues conserved between the K3L protein and the PKR substrate eIF2alpha. Mol. Cell. Biol. *17*, 4146–4158.

Kim, Y., and Orr, H.A. (2005). Adaptation in sexuals vs. asexuals: clonal interference and the Fisher-Muller model. Genetics *171*, 1377–1386.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Liao, D. (1999). Concerted evolution: molecular mechanism and biological implications. Am. J. Hum. Genet. *64*, 24–30.

Loman, N.J. and Quinlan, A.R. (2014). Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics *30*, 3399-3401.

McDonald, M.J., Rice, D.P., and Desai, M.M. (2016). Sex speeds adaptation by altering the dynamics of molecular evolution. Nature *531*, 233-236.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A.,

Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

McLysaght, A., Baldi, P.F., and Gaut, B.S. (2003). Extensive gene gain associated with adaptive evolution of poxviruses. Proc. Natl. Acad. Sci. U.S.A. *100*, 15655-15660.

Merchlinsky, M. (1989). Intramolecular homologous recombination in cells infected with temperature-sensitive mutants of vaccinia virus. J. Virol. *63*, 2030–2035.

Nei, M. (1969). Gene duplication and nucleotide substitution in evolution. Nature *221*, 40-42.

Ohno, S. (1970). In Evolution by Gene Duplication. (Springer-Verlag).

Sanjuán, R., Nebot, M.R., Chirico, N., Mansky, L.M., and Belshaw, R. (2010). Viral mutation rates. J. Virol. *84*, 9733–9748.

Searles, R.P., Bergguam, E.P., Axthelm, M.K., and Wong, S.W. (1999). Sequence and genomic analysis of Rhesus macaque rhadinovirus with similarity to Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8. J. Virol. *73*, 3040-3053.

Smith, G.L., and Moss, B. (1983). Infectious poxvirus vectors have capacity for at least 25 000 base pairs of foreign DNA. Gene *25*, 21–28.

Spyropoulos, D.D., Roberts, B.E., Panicali, D.L., and Cohen, L.K. (1988). Delineation of the viral products of recombination in vaccinia virus-infected cells. J. Virol. *62*, 1046–1054.

Stephens, S.G. (1951). Possible significance of duplication in evolution. Adv. Genet. *4*, 247-265.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., and DePristo, M.A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics *43*, 11.10.1–11.10.33.

Willer, D.O., Mann, M.J., Zhang, W., and Evans, D.H. (1999). Vaccinia virus DNA polymerase promotes DNA pairing and strand-transfer reactions. Virology *257*, 511–523.

Xiao, Y., Rouzine, I.M., Bianco, S., Acevedo, A., Goldstein, E.F., Farkov, M., Brodsky, L., and Andino, R. (2016). RNA recombination enhances adaptability and is required for virus spread and virulence. Cell Host Microbe *19*, 493–503.

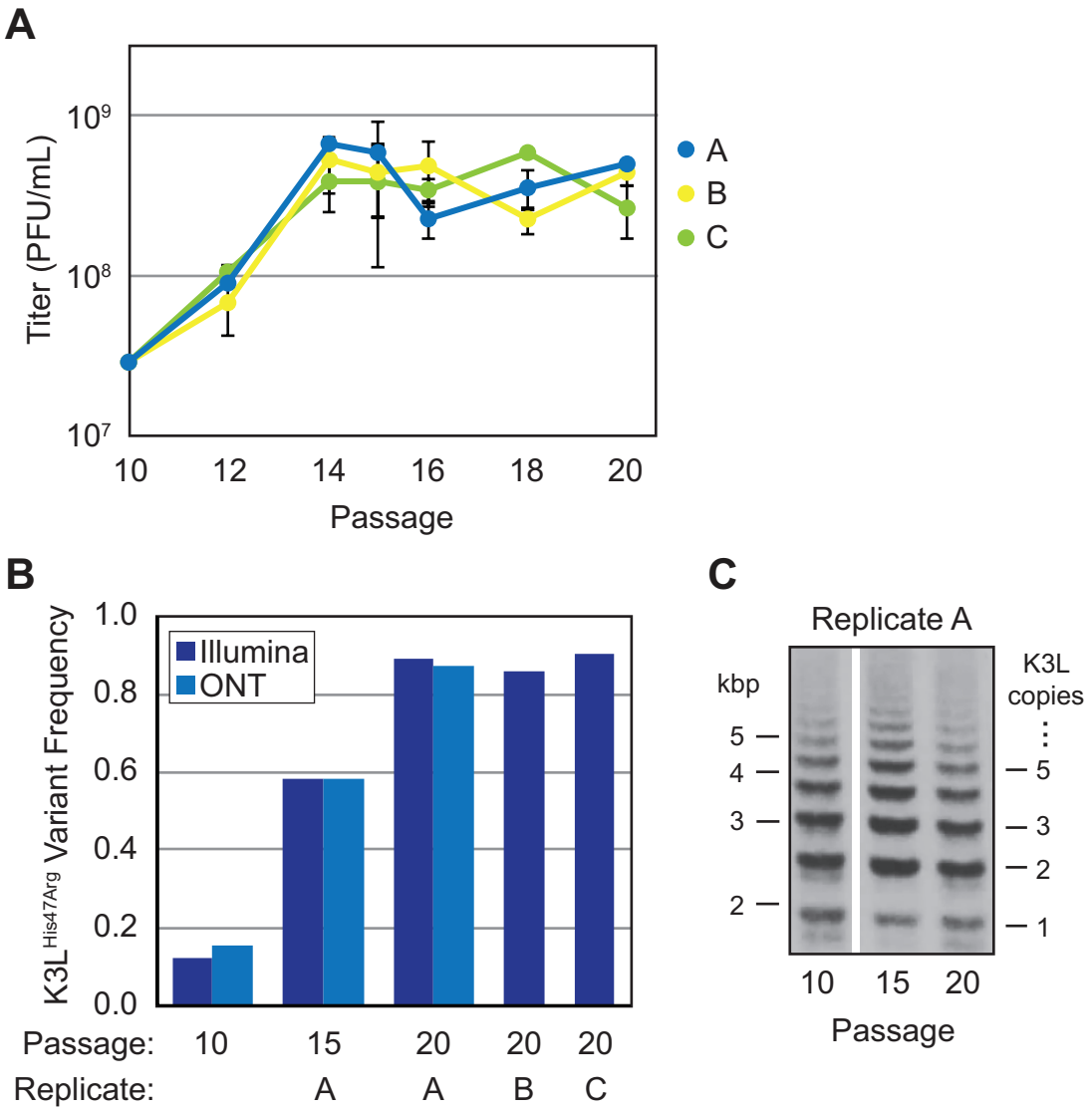


Figure 4.1. A single nucleotide variant accumulates concurrently with K3L copy number variation. (A) For each passage, HeLa cells were infected (MOI 0.1) for 48 hours. Titers were measured in BHK cells by plaque assay in triplicate, as mean PFU/mL ± standard deviation. (B) Variant frequencies were obtained by deep sequencing of viral genomic DNA on either the Illumina MiSeq or ONT MinION platform. (C) Digested viral DNA was probed with a K3L-specific probe by Southern blot analysis. Size in kbp (left) and number of K3L copies (right) are shown.

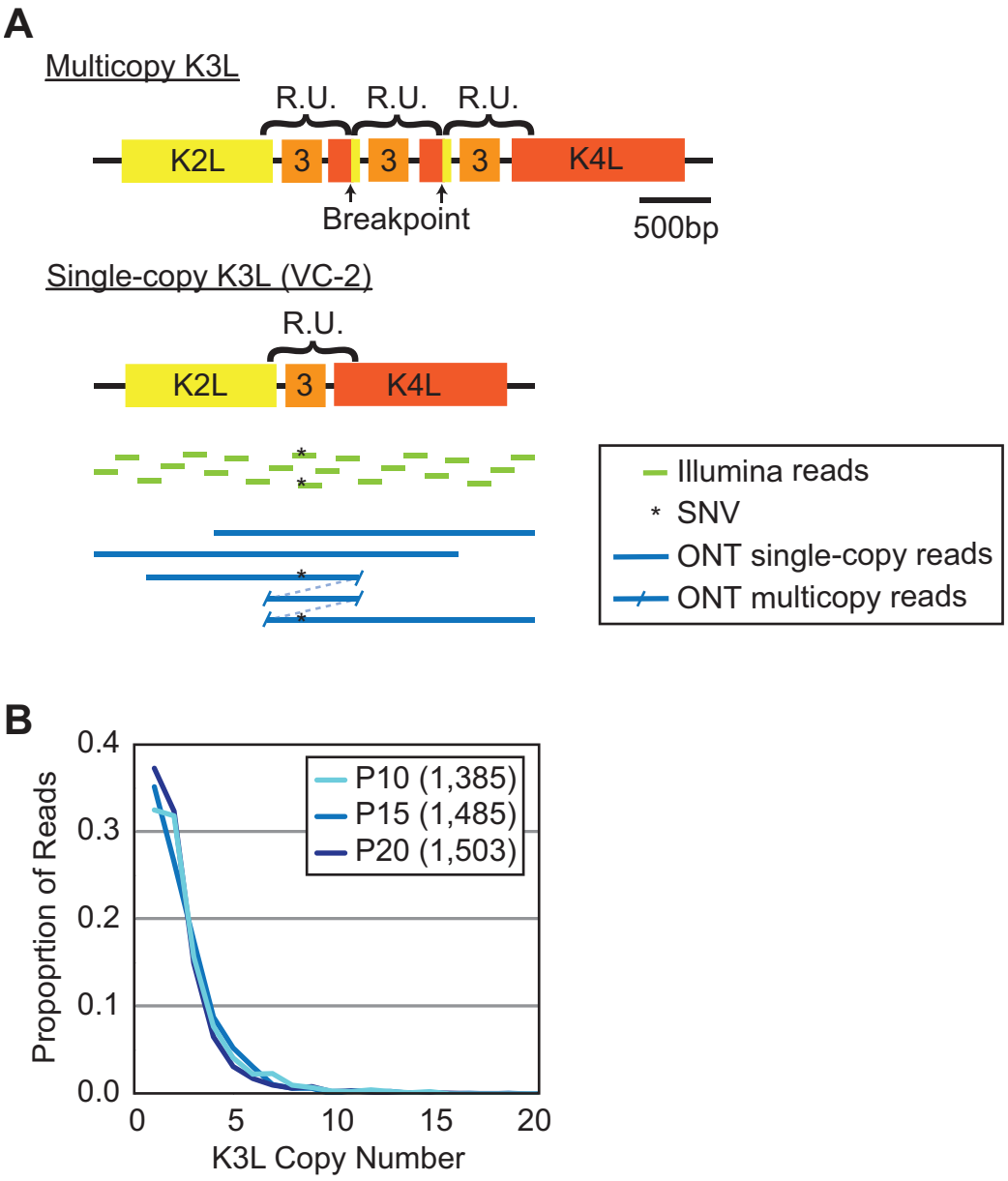
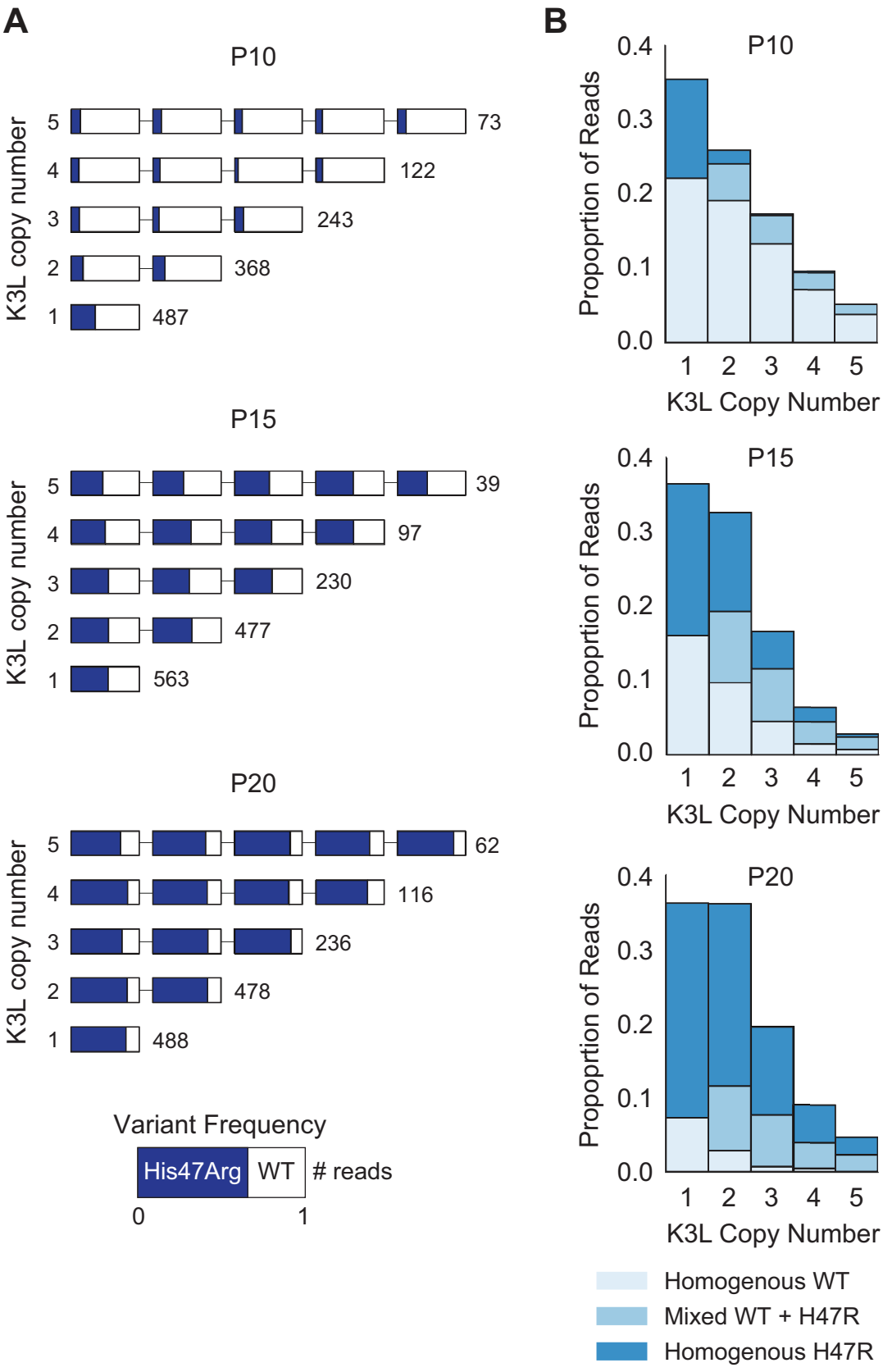


Figure 4.2. K3L^{His47Arg} SNV presence within genomes is independent of K3L copy number. (A) Representative genome structures of single-copy and multicopy K3L loci are shown on top, with the repetitive unit (R.U.) containing K3L and breakpoints indicated on each. Representative Illumina MiSeq and ONT MinION reads are shown to scale below. ONT reads are classified as single-copy or multicopy depending on how they align to the VC-2 reference genome. (B) P10, P15, and P20 viral genomes were sequenced using ONT MinION, and single-copy and multicopy reads (ONT-K3L CNV read set) were selected and analyzed for the number of K3L copies within each read. The total reads analyzed for each population are indicated in parentheses.

Figure 4.3. Concerted evolution of the K3L^{His47Arg} SNV within K3L gene arrays. (A) The ONT-K3L CNV read sets for P10, P15, and P20 were analyzed for the presence of the K3L^{His47Arg} SNV in each K3L copy. Reads containing 1-5 K3L copies oriented 5' to 3' relative to the VC-2 reference sequence are shown, and the K3L^{His47Arg} variant frequency for each copy is indicated in blue. Total reads for a given copy number are indicated to the right of each set of boxes. (B) Stacked bar plots representing the diversity of allele combinations within single-copy and multicopy reads were generated from the ONT-K3L CNV read sets. The proportion of reads containing homogenous WT, homogenous K3L^{His47Arg}, or any combinations of mixed alleles is shown for reads containing 1-5 K3L copies.


A
P10
K3L copy number
5
4
3
2
1
73
122
243
368
487
P15
39
97
230
477
563
P20
62
116
236
478
488
Variant Frequency
His47Arg
WT
reads
0
1
B
P10
Propoprtion of Reads
0.4
0.3
0.2
0.1
0.0
1 2 3 4 5
K3L Copy Number
P15
P20
Homogenous WT
Mixed WT + H47R
Homogenous H47R

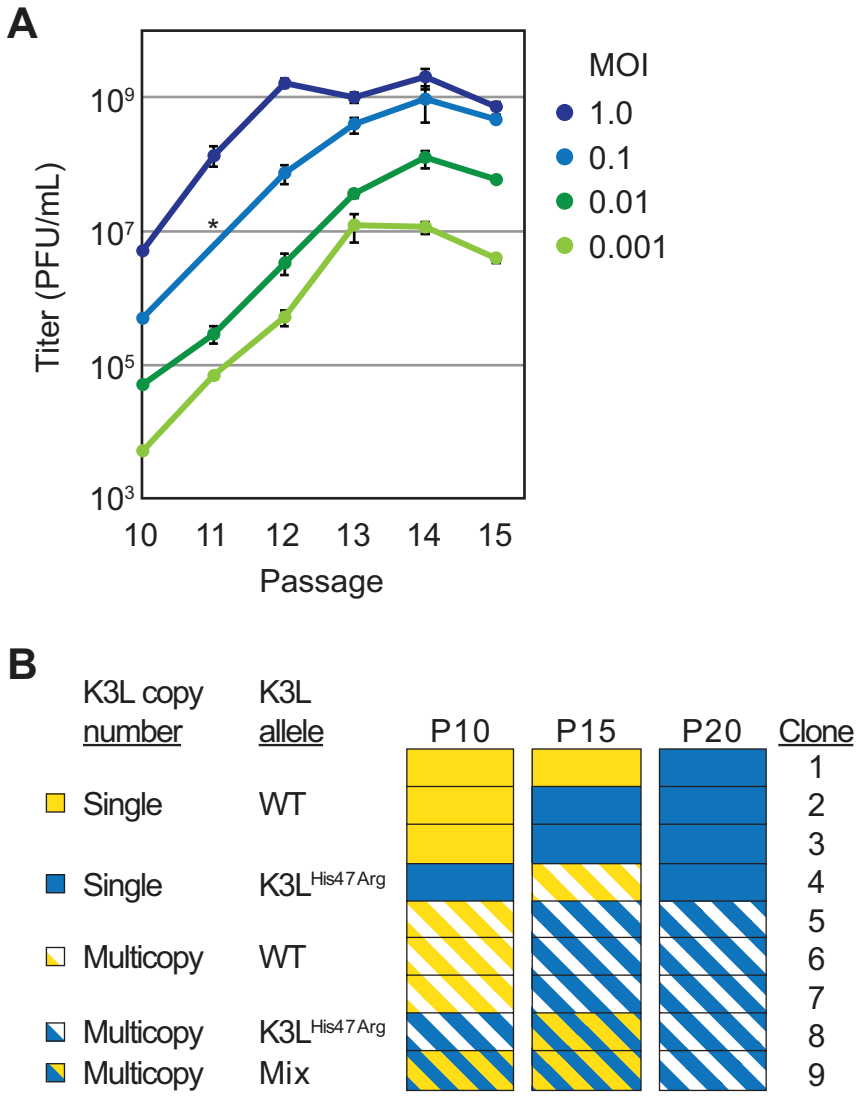


Figure 4.4. Recombination driving concerted evolution changes adaptive outcomes. (A) For each passage, HeLa cells were infected at the indicated MOI for 48 hours. Titers were measured in BHK cells by plaque assay in triplicate, as mean PFU/mL ± standard deviation. * Titer of P11 (MOI 0.1) was not assessed for technical reasons. (B) Nine viral clones from each population were isolated by four rounds of plaque purification in BHK cells. K3L copy number and presence of the K3L^{His47Arg} variant were assessed by PCR and Sanger sequencing.

Table 4.1. ONT and Illumina sequencing of replicate A populations.

	ONT			Illumina		
	P10	P15	P20	P10	P15	P20
Total reads	107,908	111,857	124,425	4,156,723	5,276,651	4,120,685
Mean read length (bp)	3,478	4,294	3,809	150	150	150
Reads aligned to VC-2	41,927	41,806	40,909	2,041,649	1,093,745	1,212,555
Reads aligned to K3L R.U. ± 75bp	1,385	1,485	1,503	n/a	n/a	n/a

CHAPTER 5

DISCUSSION

Viruses are intertwined with all cellular life in an ever-changing relationship between pathogens and their hosts. In the introduction to this dissertation, I described some of the myriad ways in which host-pathogen interactions have shaped genome evolution. Evolutionary strategies for adaptation exist in various forms, and differ between systems. Viruses are no exception, and have adapted to produce an incredibly diverse range of reproductive strategies to propagate and evade host defenses. The high degree of diversity creates challenges for fighting viral diseases, which pose a constantly changing threat to human and animal health. Thus, gaining an understanding of the inner-workings of different virus types and how they evolve is of critical importance for combatting viral infections. Poxviruses represent a virus family with a long history of causing severe disease in humans and animals, with the potential for future epidemics. Yet surprisingly little is known about how poxviruses evolve. The work in this dissertation provides new insights into the mechanisms by which poxviruses adapt, and validates experimental evolution studies in vaccinia virus as a means to discover new aspects of poxvirus biology.

One theme that emerges from this work is the significant role recombination plays in promoting poxvirus evolution. Poxviruses represent a tremendously successful virus

lineage, yet with a lower point mutation rate compared to RNA viruses (Duffy et al., 2008; Sanjuán et al., 2010), it was not previously known how these large DNA viruses adapt. Our results in Chapter 2 demonstrate a role for recombination-mediated gene amplification events in promoting the rapid adaptation of VACV populations in a genomic accordion model (Figure 5.1). The striking increase in gene copy number illustrates the inherent flexibility of poxvirus genomes, and was discovered by tracking evolving virus populations over successive generations. These copy number variation events are predicted to be transient, providing a foothold in non-optimal conditions until additional adaptive changes can occur and be selected in the population. Consistent with this idea, we observed the accumulation of beneficial point mutations following CNV at both intragenic (Figure 2.1) and extragenic (Figure 3.1) locations in VACV genomes, and a subsequent collapse of variant-containing gene arrays (Figure 2.1). Similar patterns have been observed under different selective conditions, in which gene amplifications in VACV promoting host switching (Brennan et al., 2014) were followed by the rise of adaptive point mutations and a collapse of CNV (Brennan et al., 2015). Together, these results provide evidence that structural variants generated by recombination can quickly enhance viral fitness allowing and even promoting point mutations. Additionally, the identification of adaptive gene amplification events in multiple, independent experiments supports recombination-mediated copy number variation as a common mechanism for poxvirus evolution.

Studies of variant dynamics in poxviruses have been limited, due to large genome sizes and the added complexity of recombination effects. Our studies demonstrating genetic adaptations in the form of both structural variants and single nucleotide variants

uncovers a complex system, and the work in this dissertation reveals new insights into the effects of recombination on variant dynamics in poxviruses. In Chapter 3, we found that the presence of CNV facilitates the sweep of a point mutation conferring fitness trade-offs, consistent with structural variation promoting the accumulation of mutations that might not otherwise reach fixation. To define the influence of CNV on variant dynamics more clearly, we applied new technology with the potential to reveal details of large and complex genetic regions. The ability to phase a SNV within a multicopy gene locus allowed us to discover patterns consistent with concerted evolution in VACV genomes in Chapter 4 (Figure 4.1), which is the first example of this conserved evolutionary process in a viral system. It remains to be determined whether gene conversion leading to concerted evolution is a common feature of VACV genome evolution, or whether there are other adaptive routes for the accumulation of variants in these virus populations. Nonetheless, our analyses highlight the utility of long read sequencing to uncover new aspects of complex virus population dynamics, with particular application to advance evolutionary studies of large DNA viruses.

Together, these studies validate VACV as a model system in which to discover new elements of poxvirus biology, and to explore complex evolutionary processes with broad applications. As with any scientific endeavor, our findings provoke additional questions into the complex adaptive mechanisms poxviruses employ, and the VACV experimental evolution model can be used in future studies for continued exploration. While we observed copy number variation in each of the evolving VACV populations we analyzed, there are differences in both the structural variant breakpoints generating tandem gene duplications and the suite of potentially beneficial point mutations. This

suggests that additional passaging experiments could continue to reveal new adaptive strategies. Furthermore, we have only begun to explore potential benefits afforded by the point mutations we identified in VACV genomes following selection. Our analyses of A24R variants in Chapter 3 suggest complicated mechanisms leading to fitness gains. A detailed investigation of viral transcripts could help elucidate the adaptive mechanism, and potentially uncover new aspects of transcription control (Appendix B). The other high frequency point mutations we identified in the F10L and E9L genes remain to be tested for their individual contributions to viral fitness. The E9L gene, encoding the viral DNA polymerase, is of particular interest as the vDNAP exhibits recombinase activity (Colinas et al., 1990; Hamilton and Evans, 2005), and the same point mutation was identified in two independent VACV populations. If the variant affects recombination, this could have implications for understanding how the vDNAP mediates recombination and any downstream effects on adaptation through gene amplification. There is currently no way to alter viral recombination rates without also affecting replication, and thus it is difficult to test the effects of recombination on rates of adaptation or genome evolution.

The insights we gained through *in vitro* evolution experiments with VACV provide a framework to begin to understand the critical role of recombination during poxvirus infection. Our cell culture system allows for rapid evolutionary studies, but it is unknown whether recombination and gene amplification play a similar critical role in promoting the adaptation of VACV during *in vivo* infections, which could be performed using existing mouse models of infection. Such studies could also have implications towards understanding how other, more virulent poxviruses evolve during systemic infections. We predict that other poxviruses utilize the same mechanisms of adaptation

through recombination, but no studies have yet been performed with other poxviruses. Additionally, it remains to be determined whether other large DNA viruses use similar recombination-based mechanisms to promote rapid evolution, since virus families such as *Herpesviridae* also possess slower point mutation rates compared to other types of viruses (Sanjuán et al., 2010). Giant viruses have the capacity to contain large amounts of genetic material, and represent another type of viruses with the potential to utilize gene amplification as an adaptive strategy. Thus, our findings highlighting the importance of recombination in VACV adaptation beg the question of whether this is a common strategy among other virus types.

Viruses are ubiquitous, and present a constant yet varied threat to human and animal health. An understanding of the adaptive mechanisms driving virus evolution is paramount to appropriately respond and defend against viral infections. Collectively, the work herein adds significant contributions to virus evolutionary biology by uncovering and investigating the role of recombination in promoting the rapid adaptation of poxviruses. With respect to ongoing evolutionary host-pathogen conflicts, our results contribute to the understanding of different adaptive systems, which we hope will inform future investigation of the endless coevolution between viruses and hosts.

References

Brennan, G., Kitzman, J.O., Rothenburg, S., Shendure, J., and Geballe, A.P. (2014). Adaptive gene amplification as an intermediate step in the expansion of virus host range. PLoS Pathog. *10*, e1004002.

Brennan, G., Kitzman, J.O., Shendure, J., and Geballe, A.P. (2015). Experimental evolution identifies vaccinia virus mutations in A24R and A35R that antagonize the protein kinase R pathway and accompany collapse of an extragenic gene amplification. J. Virol. *89*, 9986–9997.

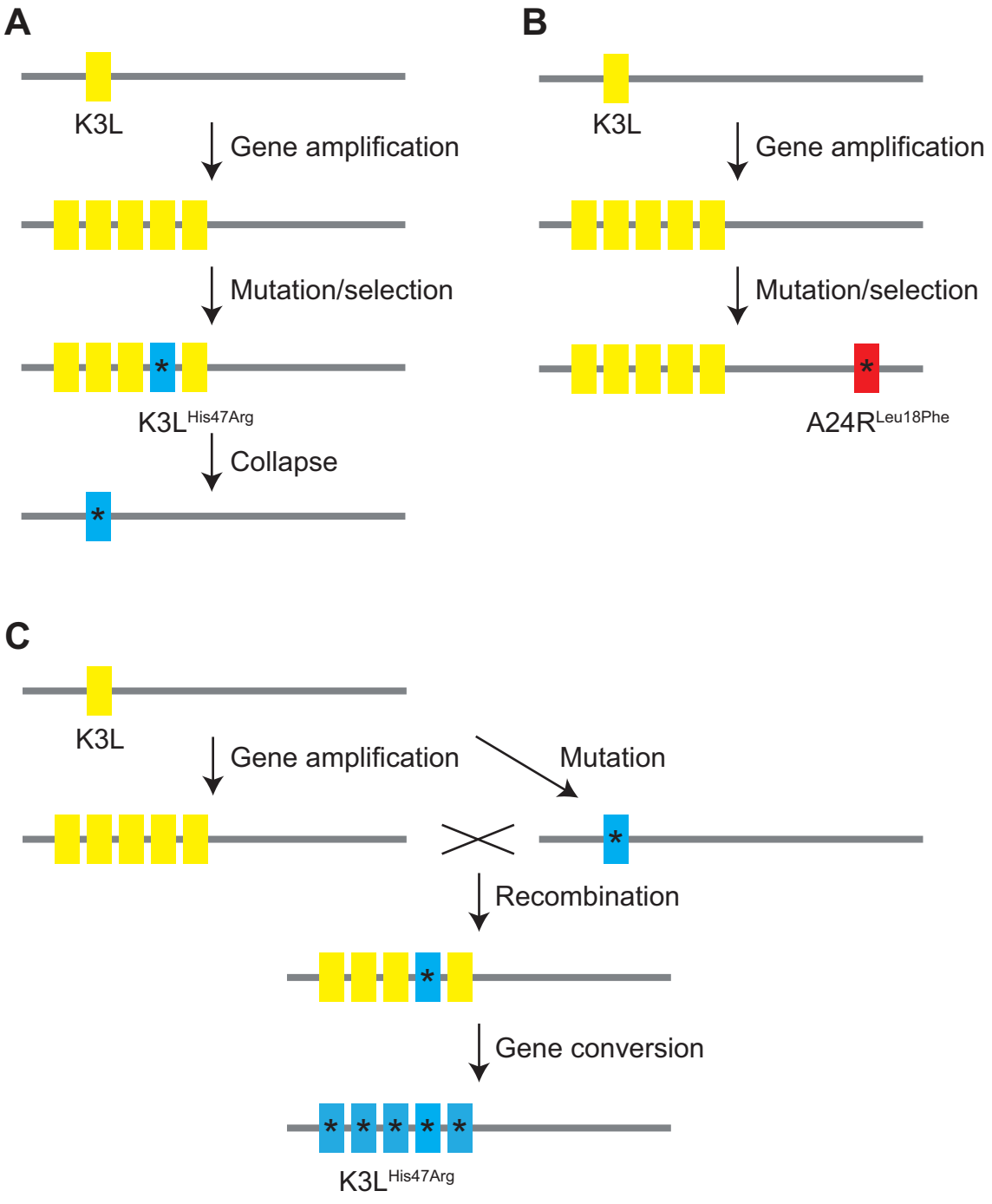
Colinas, R.J., Condit, R.C., and Paoletti, E. (1990). Extrachromosomal recombination in vaccinia-infected cells requires a functional DNA polymerase participating at a level other than DNA replication. Virus Res. *18*, 49–70.

Duffy, S., Shackelton, L.A., and Holmes, E.C. (2008). Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. *9*, 267–276.

Hamilton, M.D., and Evans, D.H. (2005). Enzymatic processing of replication and recombination intermediates by the vaccinia virus DNA polymerase. Nucleic Acids Res. *33*, 2259–2268.

Sanjuán, R., Nebot, M.R., Chirico, N., Mansky, L.M., and Belshaw, R. (2010). Viral mutation rates. J. Virol. *84*, 9733–9748.

Figure 5.1. Model of recombination-mediated adaptations in VACV genomes. (A) In the genomic accordion model, rapid gene amplification promotes viral fitness and also increases the chance of gaining beneficial intragenic point mutations. Under further selection, gene arrays can collapse back to single copies containing an advantageous point mutation. (B) Copy number variation can also promote the accumulation of extragenic point mutations inducing small fitness benefits or even fitness tradeoffs. (C) Detailed analyses of multicopy genomes alter the genomic accordion model, and suggest that gene amplification and point mutations can occur independently. The point mutation is then incorporated into multicopy genomes through recombination, and rapidly accumulates through gene conversion in a manner consistent with concerted evolution.


A
K3L
Gene amplification
Mutation/selection
K3L His47Arg
Collapse
B
K3L
Gene amplification
Mutation/selection
A24R Leu18Phe
C
K3L
Gene amplification
Mutation
Recombination
Gene conversion
K3L His47Arg

APPENDIX A

RECREATING THE EVOLUTIONARY BIRTH OF A VIRAL MIMIC

Introduction

Mimicry is a common feature found throughout nature. Visual mimics are abundant in ecological settings, such as the classic example of butterflies mimicking coloring and patterning between species (Bates, 1862; Müller, 1879; Wallace, 1865). In many such ecological cases, there is a predicted fitness advantage to defensive mimicry in evading predation (Joron and Mallet, 1998; Malcolm, 1990). These concepts can also be applied to host-pathogen interactions, since many pathogens take on features of their hosts to evade immune responses. Protein mimics in viruses are particularly prevalent, and are often identified by shared homology between viral and host defense genes. Many of these protein mimics are thought to arise via horizontal gene transfer (HGT) events followed by divergent evolution (Elde and Malik, 2009). The transfer of genetic material from the host to a virus is supported by phylogenetic analyses, in which viral genes cluster more closely with cellular hosts than with other related viruses (Rappoport and Linial, 2012). Viral protein mimics are especially abundant in poxviruses, in which up to a quarter of the genome can be composed of acquired host genes (Bugert and Darai, 2000; Hughes and Friedman, 2005).

Several protein mimics have been identified in poxviruses, which modulate host immune responses through retained homology with host factors. These include mimics of cytokines, chemokines, immune receptors, complement-regulatory factors, and anti-apoptosis factors (reviewed in Alcami, 2003; Bugert and Darai, 2000). The collective action of poxvirus protein mimics can thus evade, suppress, and prevent host immune defenses. One well-studied example of a poxvirus mimic is the vaccinia virus (VACV) factor K3L, which is a protein mimic of eukaryotic translation initiation factor 2α (eIF2α) (Beattie et al., 1991). Mammalian eIF2α is the substrate of the broadly antiviral host defense factor protein kinase R (PKR), which is activated upon infection through the detection of foreign nucleic acids (Meurs et al., 1990). Phosphorylation of eIF2α leads to a block in translation initiation, and the subsequent loss of protein production inhibits viral replication (Safer, 1983). The VACV mimic K3L relieves the translation block (Beattie et al., 1991; Carroll et al., 1993) as a competitive inhibitor of PKR, due to conserved amino acid and structural identity with eIF2α (Dar and Sicheri, 2002; Kawagishi-Kobayashi et al., 1997). VACV encodes a second PKR inhibitor, E3L, which prevents PKR activation by binding the activating nucleic acids (Chang et al., 1992; Davies et al., 1993). VACV and many other Orthopoxviruses encode both PKR inhibitors, which as host range factors likely contribute to a broad host range (Langland and Jacobs, 2002). Thus, the key role of K3L during viral infection demonstrates the potential fitness advantage afforded by a protein mimic.

Protein mimics are abundant in poxviruses, yet there are many questions surrounding the origin and evolutionary history of these viral proteins. Homology and phylogenetic evidence imply the regular exchange of genetic material from cellular

genomes to poxviruses, yet the mechanisms of HGT remain unknown. Interaction with the host genome is limited since poxviruses replicate in the cytoplasm of host cells (Moss 2013), making direct recombination events unlikely. A transposable element was discovered within the genome of taterapox (Piskurek and Okada, 2007), suggesting retrotransposition as a potential mechanism for HGT, but it remains to be seen whether host genes can be similarly inserted into poxvirus genomes through these mechanisms. Regardless of the means by which viral genomes acquire a host gene, the selection to retain the new genetic material is not well understood. Poxvirus genomes are packed with open reading frames, such that random insertion events are expected to be detrimental, or neutral at best if the insertion does not disrupt a required gene function. Finally, if a host gene is maintained, the process by which a protein mimic evolves to benefit the virus and not the host is also unknown. To explore the selection and evolution of a viral mimic, I have recreated an ancient HGT event of a host gene into a poxvirus genome. By replacing the K3L gene within VACV genomes with eIF2α, I can track adaptive changes within the viral-encoded host gene under selection to uncover adaptive processes leading to the creation of a mimic.

Results

Recreating an ancient acquisition event in VACV

To reconstruct the acquisition of the mammalian eIF2α gene by a poxvirus, I generated a series of recombinant vaccinia virus strains. In these viruses, the K3L open reading frame was replaced with different forms of the eIF2α gene (Figure A.1). The inserted host sequence is thus under the control of the native vaccinia K3L promoter to

ensure transcription from the viral genome. The recombinant viruses were generated using a ΔE3L strain (Beattie et al., 1995), producing ΔE3LΔK3L (ΔEΔK) viruses. Knocking out both PKR inhibitors placed strong selective pressure on the recombinant viruses to maintain and utilize the host gene. All recombinant viruses contain an EGFP marker used for selection, and thus I also introduced EGFP into the ΔE3L and ΔEΔK viruses as controls (Figure A.1). The different eIF2α constructs inserted into the viral genome were a full-length gene (ΔEΔK-eIF2α), an eIF2α gene encoding a S51A variant (ΔEΔK-S51A), and a C-terminal truncation of the eIF2α S51A gene generating a protein of 89 amino acids (ΔEΔK-ΔC-S51A), similar to the length of K3L (Figure A.1). These different constructs represent major differences between K3L and eIF2α, namely lack of phosphorylation by PKR and loss of the C-terminus of the protein in K3L. Either of these genetic changes could have occurred during or shortly after acquisition by the virus, and therefore the different recombinant viruses provide a means of testing whether major changes to the host gene affect adaptation within viral genomes. Additionally, this set of VACV recombinants simulates a HGT event in which host sequence has been inserted into the viral genome, to track whether the virus-encoded host gene is maintained and modified under selection.

Immediate viral fitness benefits following insertion of a host gene

Following the artificial HGT through the creation of recombinant viruses, I next determined whether the insertion of host sequence had immediate effects on viral fitness. Since all recombinant virus strains were ΔEΔK, they were generated and maintained in a support cell line that stably expresses both E3L and K3L (RK13++) (Rahman et al.,

2013). Replication comparisons of the recombinant viruses in RK13++ cells showed that there was no significant difference in titer between any of the strains tested (Figure A.1). This suggests that insertion of host sequence did not negatively impact general viral processes. However, when the viruses were propagated under selective pressure in RK13-, HeLa, or BHK cells, all ΔEΔK viruses were significantly impaired compared to the ΔE3L control (Figure A.1). This result corroborates previous findings of the requirement for K3L in different cell lines (Langland and Jacobs, 2002) classifying K3L as a host range factor. Importantly, under selection, all of the recombinant viruses containing eIF2α sequences replicated to equivalent or higher titers than the ΔEΔK virus (Figure A.1). These data suggest that VACV can tolerate and, in some instances, benefit from the acquisition of host sequence under selective conditions.

Interestingly, there were noticeable differences between replication abilities for viruses containing different forms of eIF2α. The ΔEΔK-eIF2α and ΔEΔK-S51A viruses showed no improvement over ΔEΔK; however, the ΔEΔK-ΔC-S51A virus displayed an increase in titer in all three selective cell lines (Figure A.1). This result suggests that while expressing a full-length or C-terminal truncated form of eIF2α from the virus genome does not provide any immediate fitness benefit, there is an advantage for viruses encoding the S51A variant of eIF2α. The S51A variant is not phosphorylated by PKR, yet should still be able to bind the ribosome and participate in translation initiation (Hinnebusch, 1994). Thus, by producing excess eIF2α protein which cannot be regulated by PKR, the virus likely benefits from the S51A variant relieving the block to translation. Conversely, the ΔEΔK-ΔC-S51A virus does not show similar fitness gains, despite also containing the S51A variant (Figure A.1). However, since the C-terminus of eIF2α

contains the domain needed for complex formation and interaction with the ribosome (Safer, 1983), the ΔC-S51A form of eIF2α cannot initiate translation. Viral-produced eIF2α, even the ΔC-S51A form, could still act as a competitive inhibitor of PKR, allowing endogenous eIF2α to remain active. However, as the ΔEΔK-S51A virus is the only strain with detectable fitness gains, this suggests simple overexpression may not be enough to affect PKR activity to an appreciable level. Additionally, while the S51A variant provides a fitness benefit, the observed increase in titer remains well below ΔE3L levels (Figure A.1). This implies that additional adaptations were required for optimal fitness during the creation of K3L. Yet these results suggest that a host gene in its original form can be beneficial to a virus immediately following insertion, and that by randomly acquiring host sequences, functional genes with fitness benefits may be rapidly selected in poxviruses.

Adaptive copy number variation of a host gene within viral genomes

With the set of VACV recombinants containing eIF2α genes, I applied an experimental evolution approach to follow any fitness and genetic changes that might occur over time. One of the advantages to evolutionary studies in a viral system is the ability to track genetic changes over many generations, which can reveal mechanisms underlying adaptation. To place continued selection on the virus populations to maintain and utilize the viral eIF2α gene, I performed serial passages in HeLa cells. All three eIF2α-containing recombinant virus populations quickly declined, such that after just two passages, there was not enough virus to continue the experiment (Figure A.2). Given the extremely impaired ability for these viruses to replicate in HeLa cells compared to other

cell lines (Figure A.1), this suggests that the selection pressure may have been too strong to allow the viruses to propagate. I next repeated the passaging experiment in RK13- cells, in which all of the recombinant viruses showed an intermediate replicative ability (Figure A.1). In RK13- cells, the different virus populations maintained a steady or slightly increased fitness through four passages, at which point the titer dropped sharply (Figure A.2). Thus, while the virus populations were able to persist for longer than in HeLa cells, the selection was still too great to permit sustained replication rates. I expect similar results in BHK cells, as the initial recombinant viruses reached similar titers in RK13- and BHK cells (Figure A.1). In both HeLa and RK13- cells, despite the initial fitness benefits observed with the starting ΔEΔK-S51A virus population, there were no significant differences between populations over the course of multiple passages. This suggests that small initial fitness benefits provided by the S51A variant were not sufficient to maintain an advantage under continued selective pressure.

Within the RK13- passaged viruses, there were small fitness gains in all three populations observed from passages one to four (Figure A.2). To determine whether any genetic changes had occurred in the viral eIF2α genes during these early passages, I sequenced across the eIF2α locus within viral genomes from passage four (P4) populations. This analysis revealed no detectable single nucleotide variants (SNVs) within any of the populations. However, I did detect instances in which there were multiple copies of eIF2α within individual viral genomes. Copy number variation (CNV) has been observed for viral genes in VACV (Brennan et al., 2014; Cone et al., 2017; Elde et al., 2012), but this is the first instance of a host gene undergoing copy number amplification in a viral genome. Importantly, I detected at least one example of gene

amplification in each of the different recombinant virus populations (Figure A.2). This suggests that the duplications were independent of gene size, and indeed in some cases, other viral genes were also amplified along with eIF2α (Figure A.2). These results are consistent with a pattern of random recombination events occasionally creating whole gene duplications that are then selected upon. Additional copies of viral eIF2α could benefit viral replication through overexpression to overwhelm PKR defenses, although this remains to be tested. Another advantage of CNV is the increased chance of gaining advantageous point mutations in the expanded gene, which I predict could occur if the viruses were able to continue through further passages. Even without knowing the precise benefits, the rapid accumulation of viruses with specific amplification of eIF2α suggests that recombination-mediated gene amplification following acquisition of a host gene represents a mechanism for quickly facilitating adaptation towards the creation of a mimic.

Selection for increased copy number of eIF2α within viral genomes suggests CNV provides fitness benefits, yet these virus populations suddenly declined over further passages in RK13- cells (Figure A.2). It is surprising that populations that had increased in fitness, even slightly, could so suddenly decrease. To understand how this might occur, I tested whether the decline was a direct consequence of selection by host nucleic acid sensors. For this experiment, I assessed viral titers from P4 and P5 populations after a single 48-hour infection in either RK13- cells or RK13 cells stably expressing E3L (RK13+E3L). These data show that expression of E3L in the cell line is sufficient to rescue the replication ability of the different viral populations (Figure A.2). Thus, while the P4 and P5 populations are severely hindered in RK13- cells compared to the starting

populations (Figure A.1), recovery with E3L expression suggests that the replication block stems from nucleic acid sensors like PKR. These results also indicate that the extra genetic material in P4 viruses containing eIF2α CNV is not causing the replication decline, since E3L expression is not expected to affect copy number. Overall, the passaging experiments highlight how different cellular environments can differentially affect the adaptive ability of a virus population.

Genomic context influences adaptive potential

Viral adaptive outcomes following acquisition of a host gene are certainly dependent on the environmental selective forces, but other factors may influence adaptation as well. My results showing the rescue of poorly replicating passaged populations with the expression of E3L (Figure A.2) reveal the importance of the genetic context in which a host gene is acquired. By choosing to remove E3L in addition to K3L from VACV genomes, I ensured strong selective pressure to maintain the viral copy of eIF2α. The passaging experiments showed that this selection is likely too strong to allow time for eIF2α adaptations to occur. These results, combined with the fact that E3L and K3L act in the same pathway to block PKR signaling, prompts considering whether E3L was present when eIF2α was initially introduced into poxviruses. Phylogenetic analyses of poxvirus genomes have in fact shown that E3L is likely more ancient than K3L, as E3L is shared among a wider range of poxvirus clades (Bratke et al., 2013). Thus, a more accurate reconstruction of the eIF2α HGT event should include E3L.

In consideration of the genomic context in which eIF2α was likely acquired by poxviruses, I created a new series of recombinant poxviruses, this time in the wildtype

VACV genome containing E3L. I utilized the same constructs to introduce the three different versions of eIF2α as before (Figure A.1), with the only difference being that these viruses have a functional copy of E3L (ΔK-eIF2α, ΔK-S51A, ΔK-ΔC-S51A). Replication comparisons of these ΔK3L recombinant viruses in different cell lines showed different outcomes compared to those observed with the ΔEΔK viruses. Titers were higher for ΔK3L viruses in all cell lines tested, due to the expression of E3L (Figures A.1, A.3). Consistent with their role as host range factors, E3L was sufficient to maintain wildtype levels of fitness in most cell lines, with the exception of BHKs (Figure A.3), which require K3L for optimal fitness (Langland and Jacobs, 2002). Under selection in BHK cells, the eIF2α-containing viruses displayed slight fitness defects compared to the ΔK3L control, and there was no difference between the three eIF2α constructs (Figure A.3). This suggests that the introduction of the host gene was actually detrimental under selection. However, as titers were similarly reduced in the RK13++ support cell line, yet equivalent to wildtype virus in RK13- and HeLa cells (Figure A.3), there may be some other factor affecting viral replication in these recombinant viruses. Interestingly, the ΔK-S51A virus did not show increased replication compared to the other eIF2α-containing viruses as was observed in the ΔEΔK background. Any fitness benefit afforded by the S51A variant may therefore only be uncovered under stronger selection conditions. Together these results suggest that the increased fitness afforded by E3L expression could affect the maintenance and evolution of the host gene towards a mimic.

Discussion

Collectively, these results highlight important insights towards understanding the maintenance and evolution of a protein mimic. Based on the data, I propose a new model of key adaptive steps in the creation of a viral mimic from an acquired host gene. Following HGT of eIF2α into viral genomes, changes to the PKR interface likely occurred before loss of the C-terminus of the protein. This hypothesis is supported by the finding that the S51A variant was immediately beneficial, while the ΔC-S51A form of eIF2α was not (Figure A.1). While loss of the majority of the original host protein may still have been adaptive in the evolutionary history of K3L, my results suggest that truncation likely occurred after changes to the PKR binding interface had already happened. Mutation of the serine residue in eIF2α may indeed have been a key adaptation during the creation of K3L in poxviruses, and I predict changes in PKR binding affinity could also result in an improved competitive inhibitor. This type of insight is only made possible using evolutionary experiments, and highlights the utility of an evolutionary lens to study this type of event.

Another possible key event in the creation of a mimic is adaptation through recombination. Adaptive viral gene copy number amplifications have been previously shown to promote the rapid evolution of VACV (Cone et al., 2017; Elde et al., 2012), and now I have observed the first instance of CNV of an inserted host gene (Figure A.2). Applying selection in different cell lines simulates a host-switching event, in which the viral genomic content determines the success of the population. Selection of recombination-mediated gene amplification events at the eIF2α locus suggests a benefit for viruses containing multiple copies of eIF2α, perhaps through increased protein

expression to inhibit PKR activation in a new host. Further passages are needed to explore the continued evolution of the viral-encoded host gene, and while the selective conditions I applied in the initial experiments were too strong to allow sustained viral populations, the creation of new ΔK3L recombinant viruses may provide a way forward (Figure A.3). While these experiments provide a small window into the early stages of mimic evolution, continued observations could allow us to understand how viral protein mimics evolve.

Materials and methods

Cells and viruses

RK13-, RK13+E3L, and RK13++ cells were a gift from Stefan Rothenburg (UC Davis) (Rahman et al., 2013). All RK13-derived cell lines were maintained in minimum essential medium, alpha modification (MEM-alpha; HyClone), and HeLa and BHK cells were maintained in Dulbecco's modified Eagle's medium (DMEM; HyClone), each supplemented with 10% fetal bovine serum (HyClone), 1% penicillin-streptomycin (GE Lifesciences), and 1% stable L-glutamine (GE Lifesciences). The Copenhagen strain of vaccinia virus (VC-2, WT) and the E3L deletion virus (ΔE3L) (Beattie et al., 1995) were a gift from Bertram Jacobs (Arizona State University).

Plasmid constructs

We generated a unique multiple cloning site into pBluescriptIIKS (-) (Addgene) using primers pB.2_MCS_F (5'-GAGCTCGTTTAAACGCTAGCCATATGGGTACC) and pB.2_MCS_R (5'-GGTACC CATATCGCTAGCGTTTAAACGAGCTC). Annealed

primers were digested with SacI and KpnI (New England Biolabs) and cloned into pBluescriptIIKS (-) cut with the same enzymes to generate pB.2. Sequences flanking K3L from VACV were amplified from VC-2 viral DNA: 680bp of 5' homologous sequence was amplified with the primers K2Lflank_F (5'-CTTCTTATCGATTTTTTAT ACCGAACATAAAAATAAGGTTAATTA) and K2Lflank_R (5'-CTTCTTCATATGG TGATTGTATTTCCTTGCAATTTAG), and 1024bp of 3' homologous sequence (including the native K3L promoter) was amplified with the primers K4Lflank_F (5'-CGTCGTGCGGCCGCCTTGTTAACGGGCTCGTAAATT) and K4Lflank_R (5'-CGA GCGGAGCTCGTACGATACATAGATATTACAAATATCCTAG). A VACV synthetic early/late promoter (SLP) (Chakrabarti et al., 1997) was created by annealing primers SLP_F (5' TCGACAATTGGATCAGCTTTTTTTTTTTTTTTTTGGCATATAAATA AGAAGCTTCCCGGGTCTAGAC) and SLP_R (5'-AGCTCAGATCTGGGCCCTTCG AAGAATAAATATACGGTTTTTTTTTTTTTTTTTCGACTAGGTTAAC). EGFP was amplified from pN1-EGFP (Addgene) using primers EGFP_F (5'-GGAGGACTCGAGATGGTGAGCAAGGGCGA) and EGFP_R (5'-GGAGGTATCGA TTTACTTGTACAGCTCGTCCATGC). Each PCR product was digested with restriction enzymes (New England Biolabs), gel purified (Zymo Research), and sequentially cloned into pB.2 as follows: 5' flank with ClaI and NdeI, 3' flank with NotI and SacI, SLP with SalI and XhoI, and EGFP with XhoI and ClaI. The resulting plasmid contained EGFP following SLP, between the two K3L flanking sequences (pB.2-EGFP). The K3L open reading frame was amplified from VC-2 viral DNA using primers K3L_F (5'-GTTGTAG GATCCATGCTTGCATTTTGTTATTCGTTGC) and K3L_R (5'-GTTCTTGTCGACTT ATTGATGTCTACACATCCTTTTG). The eIF2α gene was amplified from human

cDNA using primers eIF2α_F (5'-GATGTAGGATCCATGCCGGGTCTAAGTTGTAG AT) and eIF2α_R (5'-CTACTTGTCGACTTAATCTTCAGCTTTGGCTTCCAT). The resulting PCR products were cloned into pB.2-EGFP with BamHI and SalI, placing K3L or eIF2α immediately following the native K3L promoter, and upstream of SLP-EGFP to create pB.2-K3L and pB.2-eIF2α, respectively. The C-terminal truncation of eIF2α was generated by PCR from human cDNA using primers eIF2α_F and eIF2α_ΔC_R (5'-CCT CCTGTCGACTTATCTTCTT TTTGACAAATCAATATATCC). The PCR product was cloned into pB.2-EGFP, creating pB.2-eIF2α-ΔC. Site-directed mutagenesis was performed on pB.2-eIF2α or pB.2-eIF2α-ΔC using primers eIF2α_S51A_F (5'-GATAC GCCTTCTGGCTAATTCACTAAGAAGAATCATGCCTTC) and eIF2α_S51A_R (5'-GAAGGCATGATTCTTCTTAGTGAATTAGCCAGAAGGCGTATC) to generate pB.2-eIF2α-S51A and pB.2-eIF2α-ΔC-S51A, respectively.

Recombinant virus generation

Recombinant viruses were constructed by replacing the K3L gene using homologous recombination. RK13++ cells were infected with either ΔE3L or VC-2 (MOI = 1.0) and transfected at 1 hour postinfection with pB.2-EGFP, pB.2-K3L, pB.2-eIF2α, pB.2-eIF2α-S51A, or pB.2-eIF2α-ΔC-S51A plasmids by use of FuGENE6 (Promega) according to the manufacturer's protocol. Infected cells were collected at 48 hours postinfection, and viruses were released by one freeze-thaw cycle followed by sonication. Resulting viruses were plaque purified in RK13++ cells four times, selecting for recombinants expressing EGFP. Final virus clones were verified by PCR and sequencing of viral DNA across the K2L-K4L region of the genome.

Experimental evolution

For each passage, 150-mm dishes were seeded with an aliquot from the same stock of HeLa or RK13- cells (5 x 10^6 cells/dish). For P1, dishes were infected with ΔEΔK-eIF2α, ΔEΔK-S51A, or ΔEΔK-ΔC-S51A virus (MOI = 1.0) for 2 hours in a minimal volume and then supplemented with medium. After 48 hours, cells were washed, pelleted, and resuspended in 1mL of medium. Virus was released by one freeze-thaw cycle followed by sonication. 900µl of virus was then used to infect a new dish of cells for P2, and the process was repeated for subsequent passages. Viral titers were determined using the remaining 100µl of reserved virus stocks from each passage by 48-hour plaque assay in RK13++ cells. Passages were continued until the virus population was below 5 x 10^4 PFU (MOI < 0.01).

CNV and SNV analysis of viral-encoded eIF2α genes

RK13++ cells were infected with P4 ΔEΔK-eIF2α, ΔEΔK-S51A, or ΔEΔK-ΔC-S51A viruses (MOI = 0.1) for 24 hours. Virus-infected cells were collected, and total viral DNA extracted as previously described (Esposito et al., 1981). The region between K2L and K4L containing the different viral-encoded eIF2α genes was amplified by PCR with primers K2L_seq_F (5'-GGCATTGGTAAATCCTTGCAGA) and K4L_seq_R (5'-CACCTTTTAGTAGGACTAGTATCGTACAA). SNV detection was performed by sequencing across eIF2α using primers eIF2α_F and eIF2α_R for ΔEΔK-eIF2α and ΔEΔK-S51A PCR products, or eIF2α_F and eIF2α_ΔC_R (5'-CCTCCTGTCGACTTAT CTTCTTTTTGACCAATCAATATATCC) for ΔEΔK-ΔC-S51A PCR products. CNV analysis was performed by PCR using primers eIF2α_rep_F (5'-CCTCCTATGGAAGC

CAAAGCTGAAGATGAA) and eIF2α_rep_R (5'-CCTCCTATCTACAACTTAGACC CGGCAT) for ΔEΔK-eIF2α and ΔEΔK-S51A viral DNA, or eIF2α_ΔC_rep_F (5'-CCT CCTGGATATATTGATTTGTCAAAAAGAAGATAA) and eIF2α_rep_R for ΔEΔK-ΔC-S51A viral DNA. Any CNV PCR products formed were sequenced using the same primers for breakpoint detection.

References

Alcami, A. (2003). Viral mimicry of cytokines, chemokines and their receptors. Nat. Rev. Immunol. *3*, 36–50.

Bates, H.W. (1862). Contributions to an insect fauna of the Amazon valley. Trans. Linn. Soc. Lond. *23*, 495-566.

Beattie, E., Denzler, K.L., Tartaglia, J., Perkus, M.E., Paoletti, E., and Jacobs, B.L. (1995). Reversal of the interferon-sensitive phenotype of a vaccinia virus lacking E3L by expression of the reovirus S4 gene. J. Virol. *69*, 499–505.

Beattie, E., Tartaglia, J., and Paoletti, E. (1991). Vaccinia virus-encoded eIF-2 alpha homolog abrogates the antiviral effect of interferon. Virology *183*, 419–422.

Bratke, K.A., McLysaght, A., and Rothenburg, S. (2013). A survey of host range genes in poxvirus genomes. Infect. Genet. Evol. *14*, 406-425.

Brennan, G., Kitzman, J.O., Rothenburg, S., Shendure, J., and Geballe, A.P. (2014). Adaptive gene amplification as an intermediate step in the expansion of virus host range. PLoS Pathog. *10*, e1004002.

Bugert, J.J., and Darai, G. (2000). Poxvirus homologues of cellular genes. Virus Genes *21*, 111–133.

Carroll, K., Elroy-Stein, O., Moss, B., and Jagus, R. (1993). Recombinant vaccinia virus K3L gene product prevents activation of double-stranded RNA-dependent, initiation factor 2 alpha-specific protein kinase. J. Biol. Chem. *268*, 12837–12842.

Chakrabarti, S., Sisler, J.R., and Moss, B. (1997). Compact, synthetic, vaccinia virus early/late promoter for protein expression. Biotechniques *23*, 1094-1097.

Chang, H.W., Watson, J.C., and Jacobs, B.L. (1992). The E3L gene of vaccinia virus encodes an inhibitor of the interferon-induced, double-stranded RNA-dependent protein kinase. Proc. Natl. Acad. Sci. U.S.A. *89*, 4825–4829.

Cone, K.R., Kronenberg, Z.N., Yandell, M., and Elde, N.C. (2017). Emergence of a viral RNA polymerase variant during gene copy number amplification promotes rapid evolution of vaccinia virus. J. Virol. *91*, e01428-16.

Dar, A.C., and Sicheri, F. (2002). X-ray crystal structure and functional analysis of vaccinia virus K3L reveals molecular determinants for PKR subversion and substrate recognition. Mol. Cell *10*, 295–305.

Davies, M.V., Chang, H.W., Jacobs, B.L., and Kaufman, R.J. (1993). The E3L and K3L vaccinia virus gene products stimulate translation through inhibition of the double-stranded RNA-dependent protein kinase by different mechanisms. J. Virol. *67*, 1688–1692.

Elde, N.C., Child, S.J., Eickbush, M.T., Kitzman, J.O., Rogers, K.S., Shendure, J., Geballe, A.P., and Malik, H.S. (2012). Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell *150*, 831-841.

Elde, N.C., and Malik, H.S. (2009). The evolutionary conundrum of pathogen mimicry. Nat. Rev. Microbiol. *7*, 787–797.

Esposito, J., Condit, R., and Obijeski, J. (1981). The preparation of orthopoxvirus DNA. J. Virol. Methods *2*, 175–179.

Hinnebusch, A.G. (1994). The eIF-2 alpha kinases: regulators of protein synthesis in starvation and stress. Semin. Cell Biol. *5*, 417-426.

Hughes, A.L., and Friedman, R. (2005). Poxvirus genome evolution by gene gain and loss. Mol. Phylogenet. Evol. *35*, 186–195.

Joron, M., and Mallet, J.L. (1998). Diversity in mimicry: paradox or paradigm? Trends Ecol. Evol. *13*, 461–466.

Kawagishi-Kobayashi, M., Silverman, J.B., Ung, T.L., and Dever, T.E. (1997). Regulation of the protein kinase PKR by the vaccinia virus pseudosubstrate inhibitor K3L is dependent on residues conserved between the K3L protein and the PKR substrate eIF2alpha. Mol. Cell Biol. *17*, 4146-4158.

Langland, J.O., and Jacobs, B.L. (2002). The role of the PKR-inhibitory genes, E3L and K3L, in determining vaccinia virus host range. Virology *299*, 133–141.

Malcolm, S.B. (1990). Mimicry: Status of a classical evolutionary paradigm. Trends Ecol. Evol. *5*, 57–62.

Meurs, E., Chong, K., Galabru, J., Thomas, N.S., Kerr, I.M., Williams, B.R., and Hovanessian, A.G. (1990). Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon. Cell *62*, 379–390.

Moss, B. (2013). Poxvirus DNA replication. Cold Spring Harb. Perspect. Biol. *5*,

a010199.

Müller, F. (1879). *Ituna* and *Thyridia*; a remarkable case of mimicry in butterflies. Trans. Entomol. Soc. Lond. *1879*, 20-29.

Piskurek, O., and Okada, N. (2007). Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. Proc. Natl. Acad. Sci. U.S.A. *104*, 12046–12051.

Rahman, M.M., Liu, J., Chan, W.M., Rothenburg, S., and McFadden, G. (2013). Myxoma virus protein M029 is a dual function immunomodulator that inhibits PKR and also conscripts RHA/DHX9 to promote expanded host tropism and viral replication. PLoS Pathog. *9*, e1003465.

Rappoport, N., and Linial, M. (2012). Viral proteins acquired from a host converge to simplified domain architectures. PLoS Comput. Biol. *8*, e1002364.

Safer, B. (1983). 2B or not 2B: Regulation of the catalytic utilization of elF-2. Cell *33*, 7–8.

Wallace, A.R. (1865). On the phenomena of variation and geographical distribution as illustrated by the Papilionidae of the Malayan region. Trans. Linn. Soc. Lond. *25*, 19-22.

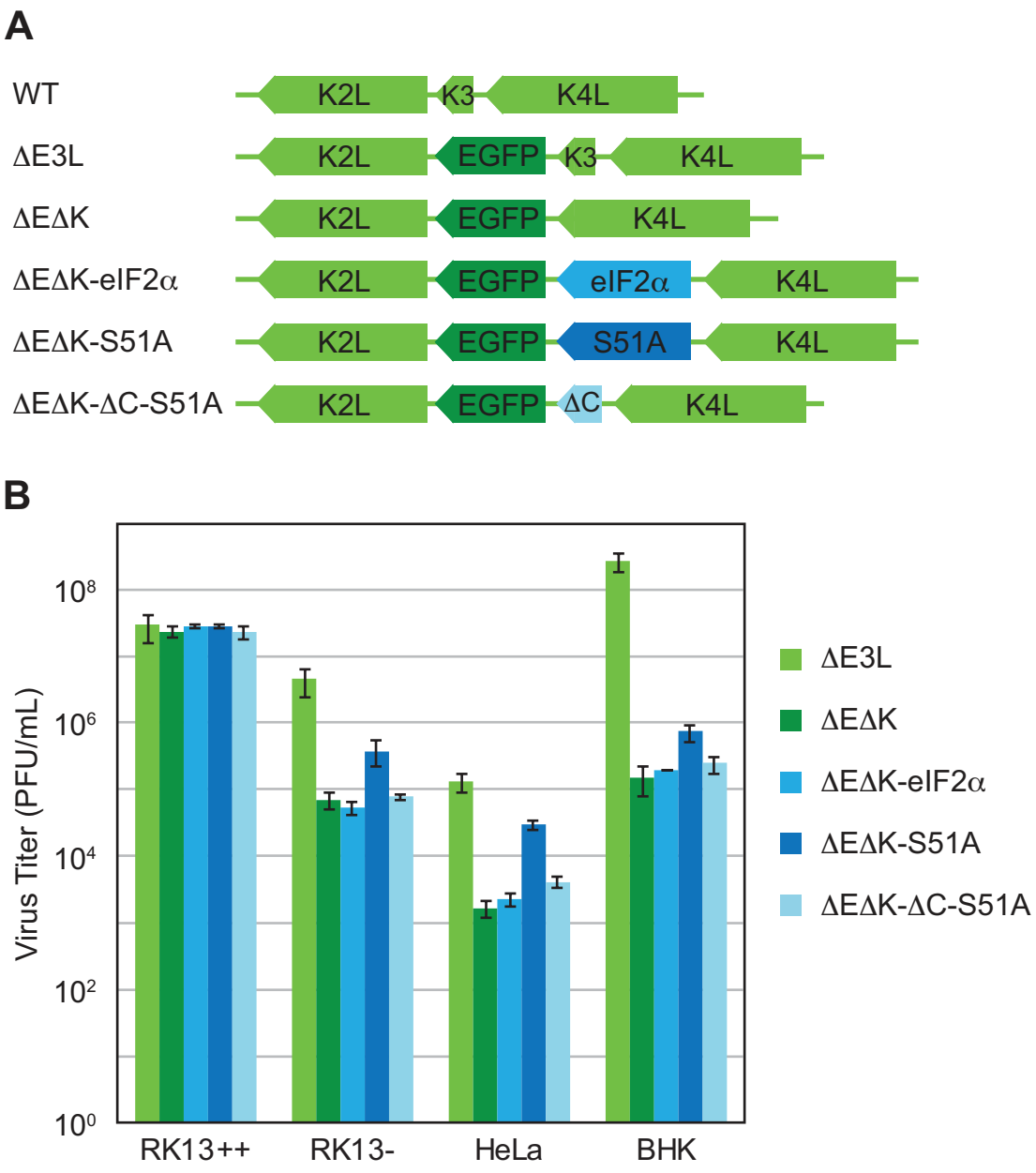
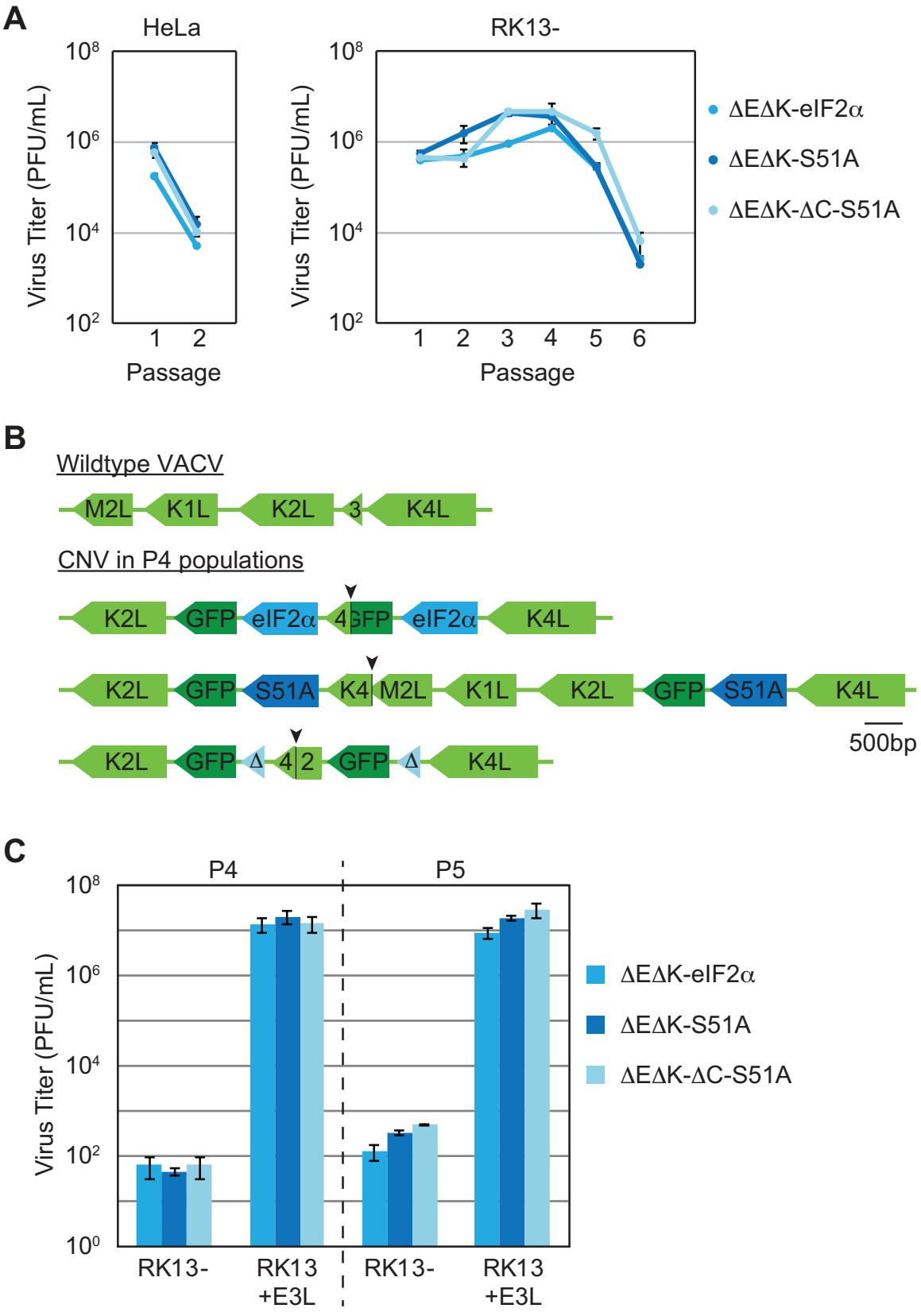


Figure A.1. Immediate fitness benefits following a simulated HGT event. (A) Genome structure of K3L (K3) region in recombinant viruses compared to wildtype (WT). (B) Cell lines were infected with the indicated recombinant viruses for 48 hours. Viral titers were measured in RK13++ cells by 48-hour plaque assay performed in triplicate as mean PFU/mL ± standard deviations.

Figure A.2. Copy number variation of a host gene in viral genomes. (A) ΔEΔK-eIF2α, ΔEΔK-S51A, or ΔEΔK-ΔC-S51A viruses were passaged in HeLa or RK13- cells until population size was too small to continue passaging. (B) Genomic structures from direct sequencing across CNV breakpoints (indicated by arrowheads above each). A representative example is shown for each P4 population compared to wildtype. (C) RK13- or RK13+E3L cells were infected with P4 or P5 viruses (MOI = 0.1) for 48 hours. All viral titers were measured in RK13++ cells by 48-hour plaque assays performed in triplicate, as mean PFU/mL ± standard deviations.


A
HeLa
RK13-
Virus Titer (PFU/mL)
10^8
10^6
10^4
10^2
1 2
1 2 3 4 5 6
Passage
ΔEΔK-eIF2α
ΔEΔK-S51A
ΔEΔK-ΔC-S51A
B
Wildtype VACV
M2L K1L K2L 3 K4L
CNV in P4 populations
K2L GFP eIF2α 4 GFP eIF2α K4L
K2L GFP S51A K4 M2L K1L K2L GFP S51A K4L
K2L GFP Δ 4 2 GFP Δ K4L
500bp
C
P4
P5
Virus Titer (PFU/mL)
10^8
10^6
10^4
10^2
10^0
RK13-
RK13 +E3L
RK13-
RK13 +E3L
ΔEΔK-eIF2α
ΔEΔK-S51A
ΔEΔK-ΔC-S51A

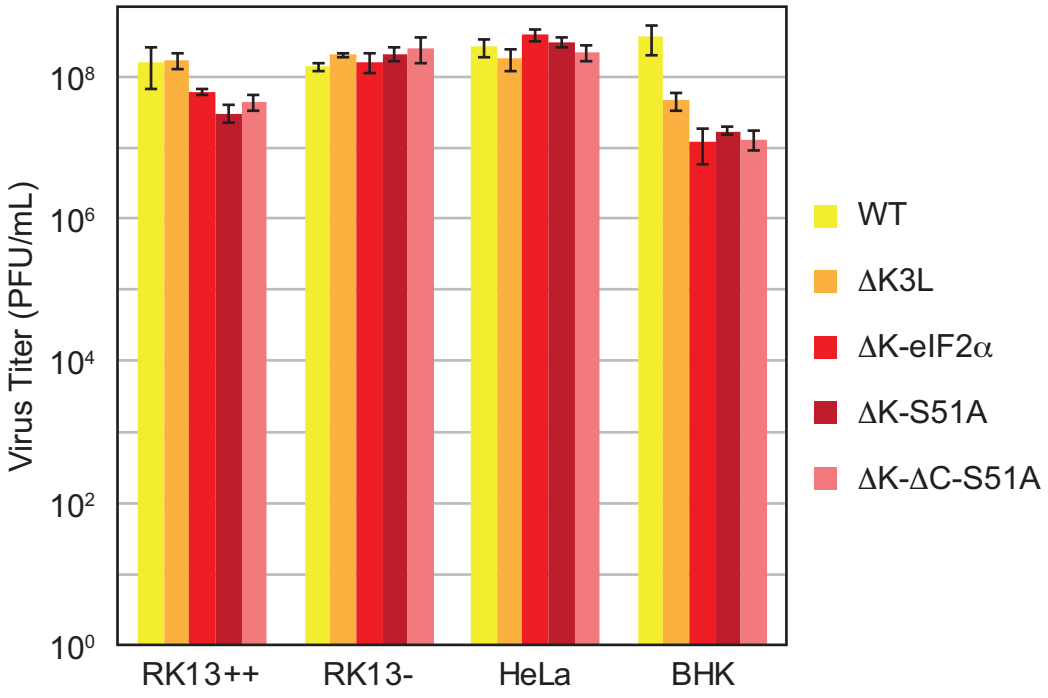


Figure A.3. HGT genomic context influences adaptive potential. (A) Cell lines were infected with the indicated recombinant viruses for 48 hours. Viral titers were measured in RK13++ cells by 48-hour plaque assay performed in triplicate as mean PFU/mL ± standard deviations.

APPENDIX B

UTILIZING A VIRAL RNA POLYMERASE VARIANT TO INVESTIGATE POXVIRUS TRANSCRIPTION

Introduction

DNA-dependent RNA polymerases (DdRPs) drive transcription in many prokaryotic and eukaryotic organisms, as well as many large DNA viruses. Evolutionary analyses have identified universal motifs in the largest polymerase subunit that are found in all DdRPs (Sonntag and Darai, 1995). Many viral polymerases share higher conservation with eukaryotic rather than prokaryotic DdRPs (Sonntag and Darai, 1995), suggesting a deep evolutionary history of similar transcription processes in DNA viruses and their hosts. The DdRP in poxviruses, for example, is predicted to have diverged with cellular homologues after the split between eukaryotes and prokaryotes (Broyles and Moss, 1986), but before the generation of eukaryotic RNA polymerase (RNAP) I, II, and III (Sonntag and Darai, 1995). Poxviruses replicate solely in the cytoplasm of host cells (Moss, 2013), and thus depend on the viral RNA polymerase (vRNAP) for successful transcription and propagation.

Cytoplasmic replication separates poxviruses from many other large DNA viruses, and poxviruses are thus less reliant on host factors for viral mRNA synthesis. The vRNAP in poxviruses is composed of nine subunits, and transcribes the set of early,

intermediate, and late genes in a temporal manner (reviewed in Broyles, 2003). Gene expression control is at the level of transcription initiation, and relies heavily on the cascade of viral transcription factors produced throughout the viral lifecycle that recognize different viral promoters (Moss, 2013). Early transcripts are produced within the viral core immediately following virion fusion, and thus rely solely on viral factors packaged in the virion (Broyles, 2003, Moss 2013). Intermediate and late gene transcription take place in the cytoplasm post replication, and depend on both viral and cellular transcription factors (Broyles, 2003; Oh and Broyles, 2005). It is not well understood how transcription factors interact with the vRNAP, or even which host factors are required for viral mRNA production. Thus, transcriptional control in poxviruses remains an active area of study.

One key difference between early and intermediate/late genes is the control of transcription termination. The vRNAP complex during early transcription includes all nine viral-encoded subunits, whereas only eight are found in the post-replicative transcription complex (Wright and Coroneos, 1995). The extra subunit is important for recognition of a specific termination signal in early genes, leading to homogenous transcript lengths (reviewed in Piacente et al., 2008). In contrast, there is no identified intermediate/late gene termination signal, although there is evidence to support an active termination mechanism (Condit and Niles, 2002). The result is extremely heterogeneous 3' ends for intermediate and late transcripts, ranging from 2-5kbp in length (Cooper et al., 1981; Mahr and Roberts, 1984; Xiang et al., 2000). The mechanisms determining post-replicative transcription termination thus remain largely unknown. Most of the current knowledge surrounding poxvirus transcription stems from studies of the model poxvirus,

vaccinia virus (VACV), and I propose to utilize a VACV vRNAP variant that I identified in Chapter 3 (Cone et al., 2017) to uncover new aspects of poxvirus transcription control.

Results

Poxvirus transcription is mediated by the viral RNA polymerase in complex with viral and cellular factors. Transcription control is critical for both the propagation of virus particles, and the evasion of host nucleic acid sensors. Double-stranded RNA (dsRNA) is produced by many different viruses (Weber et al., 2006), and activates broadly antiviral host cytoplasmic dsRNA sensors like protein kinase R (PKR) and oligoadenylate synthetase 1 (OAS-1) (reviewed in Sparrer and Gack, 2015; Wu and Chen, 2014). In VACV, dsRNA is formed by overlapping, complementary mRNA transcripts (Boone et al., 1979; Colby and Duesberg, 1969), and likely exacerbated by the long, heterogeneous 3' untranslated regions of intermediate/late transcripts (Cooper et al., 1981; Mahr and Roberts, 1984). Thus, alterations to transcription termination control could reduce dsRNA production and activation of host sensors. In Chapter 3, I identified a variant in the VACV A24R gene, encoding a subunit of the vRNAP, that arose in a ΔE3L virus population under selection to adapt to PKR sensing (Cone et al., 2017). This A24R Leu18Phe variant was sufficient for a fitness increase under selective conditions, and also reduced total dsRNA production. While there was no resulting effect on OAS-1 activity, surprisingly, the A24R variant increased activation of PKR. Thus, while the mechanism by which the variant confers an advantage is unclear, these results suggest an interesting change to viral RNA production induced by the A24R variant.

A reduction in dsRNA with a concurrent increase in PKR activation suggests that

there is likely something different about the quality, rather than the amount, of viral RNAs being produced in the presence of the A24R Leu18Phe variant. Previously identified mutations in A24R have previously been shown to reduce transcript length (Condit et al., 1991; Prins et al., 2004), which could affect total dsRNA levels. To determine whether the A24R Leu18Phe variant similarly induced changes to transcript length, we performed RNA-seq of viral transcripts. The A24R Leu18Phe variant provided a small fitness benefit under selection in human fibroblast (HF) cells; however, because it arose in a ΔE3L background, replication was still reduced compared to a wildtype virus (Cone et al., 2017). Therefore, to optimize transcript production for sequencing, we chose to collect RNA from infected HF cells in which PKR had been knocked down (HF-shPKR) (Child et al., 2012). This maintains the HF environment, while increasing replication for both the A24R^{Leu18Phe} recombinant virus and the ΔE3L control (Figure B.1). Interestingly, the fitness increase observed with the A24R^{Leu18Phe} virus in HF cells is not maintained in HF-shPKR cells (Figure B.1). This suggests that the adaptive benefit is dependent on PKR, and supports a complex mechanism of action since increased PKR activation in HF cells is not expected to aid replication. However, this result supports the need to understand the subtle changes to viral RNA that might underlie these phenotypes.

To analyze viral transcript length, we sequenced total RNA from ΔE3L or A24R^{Leu18Phe}-infected HF-shPKR cells. Sequencing was performed on the Oxford Nanopore Technologies MinION platform, which with long read sequencing capabilities has the potential to sequence an entire transcript in a single read. We obtained over 4,000 viral reads for each sample, with an average read length of about 2kbp. Comparisons of

overall transcript length from this analysis showed no significant difference between transcripts produced by the ΔE3L or A24R^{Leu18Phe} virus (Figure B.1). This suggests that while the A24R^{Leu18Phe} virus produced less total dsRNA (Cone et al., 2017), it was not a result of overall transcript length reduction. However, since it is predicted that intermediate/late genes are primarily responsible for dsRNA production (Boone et al., 1979; Colby and Duesberg, 1969), we also assessed whether different types of transcripts showed alterations in transcript length. For this analysis, we selected reads aligning to individual viral genes, and again saw no difference between the ΔE3L or A24R^{Leu18Phe} virus transcripts, regardless of whether they were from early, intermediate, or late class genes (Figure B.1). While these data suggest that there is no change to overall or individual transcript length with the A24R Leu18Phe variant, we propose that additional analyses with increased depth of sequencing could reveal differences among a subset of viral genes. Additionally, analyses of transcript abundance, which were not feasible with our initial low depth of coverage, could provide insight into whether certain transcripts are differentially regulated in the presence of the A24R variant.

Conclusions and future directions

While preliminary experiments analyzing viral transcripts produced by a vRNAP variant did not show differences in transcript length, our combined results suggest that the A24R Leu18Phe variant could be used to investigate additional aspects of poxvirus transcription. The observed reduction in total dsRNA, increase in PKR activation, but no effect on OAS-1 activity with the A24R Leu18Phe variant are consistent with a subtle change in viral transcription that effects some, but not all transcripts. Changes to

transcript termination are still possible, since our initial RNA-seq analysis did not provide complete genome coverage or enough depth to assess transcript abundance. Furthermore, the early transcripts we analyzed showed an unexpected amount of variation in length (Figure B.1), given the mechanisms regulating early gene transcription termination (Piacente et al., 2008). This suggests that at the late time point we analyzed transcripts, the reads mapping to early genes may represent read-through of intermediate/late genes or aberrant use of early promoter elements by the post-replicative transcription complex. Thus, a more in-depth analysis of viral transcripts at different times post-infection is needed to assess whether or how the A24R Leu18Phe variant affects viral transcription. We expect that such a thorough analysis could reveal subtle changes in the transcriptome that may account for the phenotypes observed.

Additional sequencing experiments could also reveal new insights into vaccinia transcription control. Oxford Nanopore Technologies recently released a direct RNA sequencing platform, which could greatly aid in analysis of viral transcripts in their original form. Added selection for different subsets of transcripts could provide a more targeted analysis of the VACV dsRNAs that activate sensors, such as using a stranded RNA preparation to identify overlapping transcripts, or isolating capped mRNAs to ensure sequencing of whole transcripts and not fragments. We also expect that pull-down experiments with a dsRNA or PKR antibody prior to RNA sequencing could reveal the dsRNAs produced during viral infection that bind and activate PKR, which is currently unknown. Finally, to stabilize viral transcripts and improve yield for sequencing, we could utilize a cell line in which the host 5'-3' exonuclease, Xrn1, is knocked down. However, recent studies showed that along with poxvirus-encoded decapping enzymes,

Xrn1 is required to reduce dsRNA levels during infection (Burgess and Mohr, 2015; Liu et al., 2015). Thus, while Xrn1 knockdown enhances the antiviral activities of PKR and OAS (Burgess and Mohr, 2015), it would be interesting to test whether the A24R Leu18Phe variant confers resistance to these conditions. We predict that it would, based on the reduction of dsRNA during A24R^{Leu18Phe} virus infection with a tolerance for increased PKR activity. However, it may require moving the A24R variant into a wildtype background in which E3L is expressed to allow sufficient virus replication. Any of these future experiments could elucidate new aspects of poxvirus transcription control, and the relationship between viral dsRNA production and activation of host sensors.

Another avenue for pursuit using RNA sequencing of viral transcripts is to investigate the role for the adenosine deaminases acting on RNA (ADARs), which use dsRNA as a substrate. ADARs edit and destabilize dsRNAs through the deamination of adenosine to inosine, and can thus affect dsRNAs produced during viral infections (reviewed in Samuel, 2011). ADARs generally reduce interferon responses through an unknown mechanism (Samuel, 2011), and one isoform of ADAR1 is interferon-induced (George et al., 2011; Toth et al., 2006), consistent with a regulatory role for ADARs in controlling interferon responses. The role for ADARs during viral infection are complex, and in fact can be antiviral or pro-viral depending on the virus and cellular host (Samuel, 2011). ADARs have been shown to directly edit viral RNAs, but it is unknown whether editing occurs on poxvirus RNAs (Samuel, 2011). ADAR1 can suppress the activation of PKR during viral infection (Nie et al., 2006; Toth et al., 2009), but interestingly, the VACV E3L protein inhibits both PKR (Chang et al., 1992) and ADAR1 (Liu et al., 2001). E3L and ADAR1 share homology in the Z-DNA binding domain (Patterson and

Samuel, 1995), which may contribute to the mechanism of inhibition by E3L. These confusing observations demonstrate the need to understand how ADAR activity affects poxvirus-produced dsRNAs and viral replication.

The abundant dsRNAs produced during poxvirus infections, and the cytoplasmic location of transcription coinciding with ADAR1 localization, makes poxvirus transcripts a likely target for editing. One method to determine whether ADARs act on poxvirus transcripts is to utilize the RNA sequencing analyses we proposed above. Identification of A-to-G (or U-to-C) changes in viral sequences would be indicative of ADAR activity acting directly on viral transcripts. To understand the role of E3L in moderating ADAR1 activity, alterations to the Z-DNA binding domain could be tested for effects on ADAR1 inhibition or direct interaction. The relationship between E3L and ADAR1 also has implications for the interpretation of our A24R variant results. The A24R Leu18Phe variant arose in a ΔE3L population, which may have exposed the virus to antagonism by PKR as well as ADAR1. Perhaps alterations to dsRNA production by the A24R Leu18Phe variant affect ADAR1 activity with the trade-off of increasing PKR activity. Thus, we would also like to test whether the A24R Leu18Phe variant provides fitness benefits through overcoming ADAR1. Determining whether the $A24R^{Leu18Phe}$ virus is affected by ADAR1 knockdown could be very informative in this respect. Clearly there are a lot of future directions to pursue in understanding the web of interactions between viral dsRNA production, host immune responses, and viral evasion techniques. New advents in sequencing technology combined with experimental evolution studies provide a means of investigation that could move the field forward.

Materials and methods

Cells

HF-shCtrl and HF-shPKR cells (Child et al., 2012) were a gift from Adam Geballe (Fred Hutchinson Cancer Research Center). HF-shCtrl, HF-shPKR, and BHK cells were maintained in Dulbecco's modified Eagle's medium (DMEM; HyClone) supplemented with 10% fetal bovine serum (HyClone), 1% penicillin-streptomycin (GE Lifesciences), and 1% stable L-glutamine (GE Lifesciences).

RNA-seq

HF-shPKR cells were infected with ΔE3L (Beattie et al., 1995) or A24R^{Leu18Phe} virus (Cone et al., 2017) (MOI = 20) for 8 hours. Total polyadenylated RNA was isolated as in (Yang et al., 2010), followed by rRNA depletion with the NEBNext® rRNA depletion kit (New England Biolabs) and mRNA selection with Dynabeads (Thermo Fisher Scientific), each according to the manufacturer's protocol. The cDNA library was prepared using a SQK-NSK007 sample prep kit (Oxford Nanopore Technologies), and sequenced on a MinION Mk1B devide using R9 chemistry MinION flow cells (Oxford Nanopore Technologies). Sequenced reads were base-called using the Metrichor cloud-based software suite (Oxford Nanopore Technologies), and FASTA and FASTQ sequences were extracted with *poretools* (Loman and Quinlan, 2014). Reads were aligned to the Copenhagen reference strain of VACV (VC-2; accession M35027.1; modified on poxvirus.org) (Goebel et al., 1990) using BWA-MEM in default mode (Li and Durbin, 2009).

References

Beattie, E., Denzler, K.L., Tartaglia, J., Perkus, M.E., Paoletti, E., and Jacobs, B.L. (1995). Reversal of the interferon-sensitive phenotype of a vaccinia virus lacking E3L by expression of the reovirus S4 gene. J. Virol. *69*, 499–505.

Boone, R.F., Parr, R.P., and Moss, B. (1979). Intermolecular duplexes formed from polyadenylylated vaccinia virus RNA. J. Virol. *30*, 365–374.

Broyles, S.S. (2003). Vaccinia virus transcription. J. Gen. Virol. *84*, 2293–2303.

Broyles, S.S., and Moss, B. (1986). Homology between RNA polymerases of poxviruses, prokaryotes, and eukaryotes: nucleotide sequence and transcriptional analysis of vaccinia virus genes encoding 147-kDa and 22-kDa subunits. Proc. Natl. Acad. Sci. U.S.A. *83*, 3141–3145.

Burgess, H.M., and Mohr, I. (2015). Cellular 5'-3' mRNA exonuclease Xrn1 controls double-stranded RNA accumulation and anti-viral responses. Cell Host Microbe *17*, 332–344.

Chang, H.W., Watson, J.C., and Jacobs, B.L. (1992). The E3L gene of vaccinia virus encodes an inhibitor of the interferon-induced, double-stranded RNA-dependent protein kinase. Proc. Natl. Acad. Sci. U.S.A. *89*, 4825–4829.

Child, S.J., Brennan, G., Braggin, J.E., and Geballe, A.P. (2012). Species specificity of protein kinase R antagonism by cytomegalovirus TRS1 genes. J. Virol. *86*, 3880-3889.

Colby, C., and Duesberg, P.H. (1969). Double-stranded RNA in vaccinia virus infected cells. Nature *222*, 940–944.

Condit, R.C., Easterly, R., Pacha, R.F., Fathi, Z., and Meis, R.J. (1991). A vaccinia virus isatin-beta-thiosemicarbazone resistance mutation maps in the viral gene encoding the 132-kDa subunit of RNA polymerase. Virology *185*, 857–861.

Condit, R.C., and Niles, E.G. (2002). Regulation of viral transcription elongation and termination during vaccinia virus infection. Biochim. Biophys. Acta *1577*, 325–336.

Cone, K.R., Kronenberg, Z.N., Yandell, M., and Elde, N.C. (2017). Emergence of a viral RNA polymerase variant during gene copy number amplification promotes rapid evolution of vaccinia virus. J. Virol. *91*, e01428-16.

Cooper, J.A., Wittek, R., and Moss, B. (1981). Extension of the transcriptional and translational map of the left end of the vaccinia virus genome to 21 kilobase pairs. J. Virol. *39*, 733–745.

George, C.X., Gan, Z., Liu, Y., and Samuel, C.E. (2011). Adenosine deaminases acting on RNA, RNA editing, and interferon action. J. Interferon Cytokine Res. *31*, 99–117.

Goebel, S.J., Johnson, G.P., Perkus, M.E., Davis, S.W., Winslow, J.P., and Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. Virology *179*, 247–66, 517–63.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Liu, S.W., Katsafanas, G.C., Liu, R., Wyatt, L.S., and Moss, B. (2015). Poxvirus decapping enzymes enhance virulence by preventing the accumulation of dsRNA and the induction of innate antiviral responses. Cell Host Microbe *17*, 320–331.

Liu, Y., Wolff, K.C., Jacobs, B.L., and Samuel, C.E. (2001). Vaccinia virus E3L interferon resistance protein inhibits the interferon-induced adenosine deaminase A-to-I editing activity. Virology *289*, 378–387.

Loman, N.J. and Quinlan, A.R. (2014). Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics *30*, 3399-3401.

Mahr, A., and Roberts, B.E. (1984). Arrangement of late RNAs transcribed from a 7.1-kilobase EcoRI vaccinia virus DNA fragment. J. Virol. *49*, 510–520.

Nie, Y., Hammond, G.L., and Yang, J.H. (2007). Double-stranded RNA deaminase ADAR1 increases host susceptibility to virus infection. J. Virol. *81*, 917–923.

Oh, J., and Broyles, S.S. (2005). Host cell nuclear proteins are recruited to cytoplasmic vaccinia virus replication complexes. J. Virol. *79*, 12852–12860.

Patterson, J.B., and Samuel, C.E. (1995). Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. Mol. Cell. Biol. *15*, 5376–5388.

Piacente, S., Christen, L., Dickerman, B., Mohamed, M.R., and Niles, E.G. (2008). Determinants of vaccinia virus early gene transcription termination. Virology *376*, 211–224.

Prins, C., Cresawn, S.G., and Condit, R.C. (2004). An isatin-beta-thiosemicarbazone-resistant vaccinia virus containing a mutation in the second largest subunit of the viral RNA polymerase is defective in transcription elongation. J. Biol. Chem. *279*, 44858–44871.

Samuel, C.E. (2012). ADARs: viruses and innate immunity. Curr. Top. Microbiol. Immunol. *353*, 163–195.

Sonntag, K.C., and Darai, G. (1995). Evolution of viral DNA-dependent RNA polymerases. Virus Genes *11*, 271–284.

Sparrer, K.M., and Gack, M.U. (2015). Intracellular detection of viral nucleic acids. Curr. Opin. Microbiol. *26*, 1–9.

Toth, A.M., Li, Z., Cattaneo, R., and Samuel, C.E. (2009). RNA-specific adenosine deaminase ADAR1 suppresses measles virus-induced apoptosis and activation of protein kinase PKR. J. Biol. Chem. *284*, 29350–29356.

Toth, A.M., Zhang, P., Das, S., George, C.X., and Samuel, C.E. (2006). Interferon action and the double-stranded RNA-dependent enzymes ADAR1 adenosine deaminase and PKR protein kinase. Prog. Nucleic Acid Res. Mol. Biol. *81*, 369–434.

Weber, F., Wagner, V., Rasmussen, S.B., Hartmann, R., and Paludan, S.R. (2006). Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. J. Virol. *80*, 5059–5064.

Wright, C.F., and Coroneos, A.M. (1995). The H4 subunit of vaccinia virus RNA polymerase is not required for transcription initiation at a viral late promoter. J. Virol. *69*, 2602–2604.

Wu, J., and Chen, Z.J. (2014). Innate immune sensing and signaling of cytosolic nucleic acids. Annu. Rev. Immunol. *32*, 461–488.

Xiang, Y., Latner, D.R., Niles, E.G., and Condit, R.C. (2000). Transcription elongation activity of the vaccinia virus J3 protein in vivo is independent of poly(A) polymerase stimulation. Virology *269*, 356–369.

Yang, Z., Bruno, D.P., Martens, C.A., Porcella, S.F., and Moss, B. (2010). Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing. Proc. Natl. Acad. Sci. U.S.A. *107*, 11513–11518.

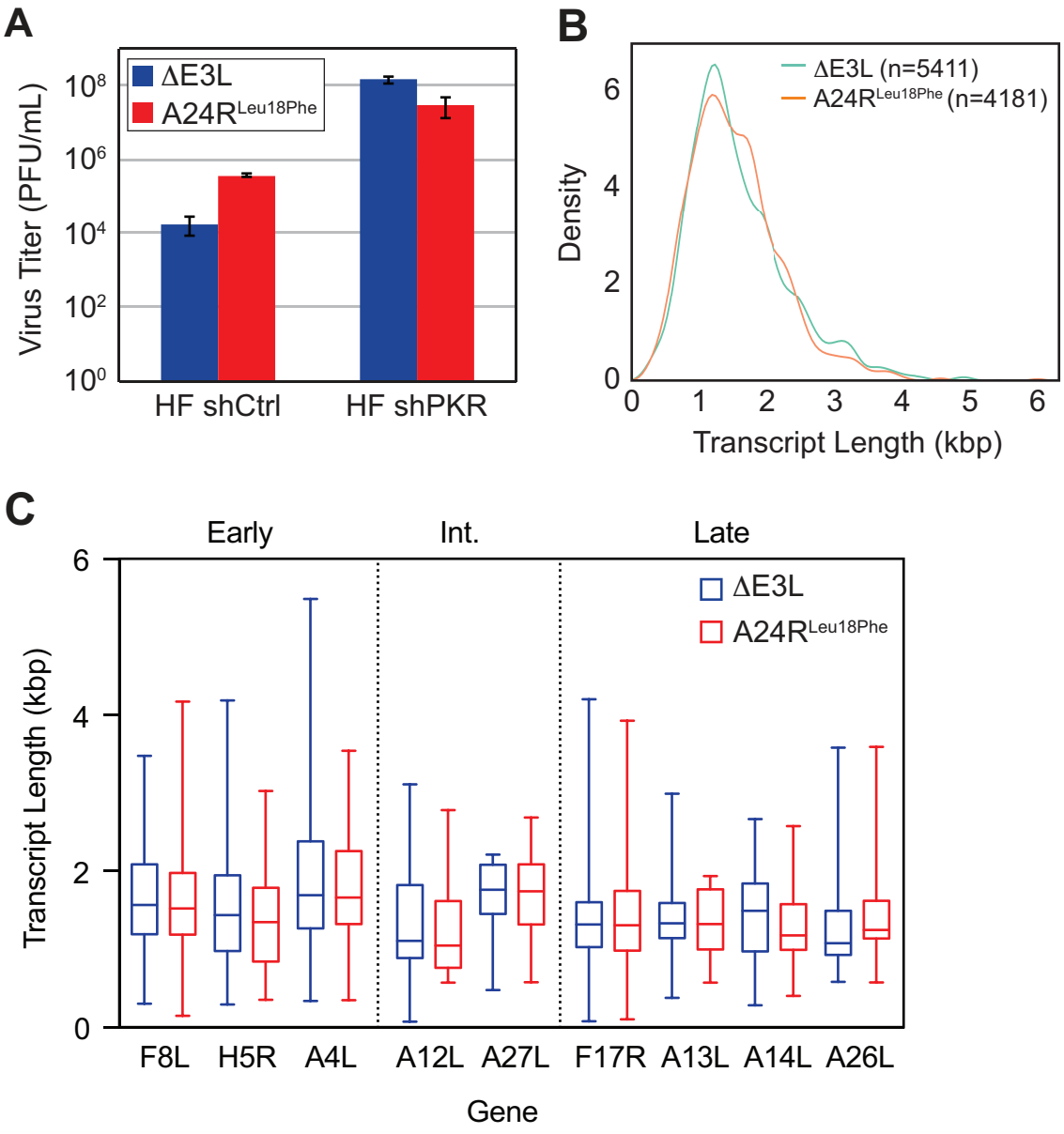


Figure B.1. The A24R Leu18Phe variant does not affect transcript length. (A) HF-shCtrl or HF-shPKR cells were infected with the ΔE3L or A24R^{Leu18Phe} virus (MOI = 0.1) for 48 hours. Viral titers were measured in BHK cells by 72-hour plaque assay performed in triplicate, as mean PFU/mL ± standard deviations. (B) HF-shPKR cells were infected with ΔE3L or A24R^{Leu18Phe} virus (MOI = 20) for 8 hours. Total polyadenylated RNA was isolated, and cDNA prepared and sequenced on the MinION platform. Viral reads were analyzed for transcript length, and the number of reads is shown in parentheses. (C) Viral genes with at least 20 reads spanning the open reading frame and any downstream sequences were analyzed for transcript length. Data are shown as box-and-whisker plots, and there is no significant difference between any ΔE3L or A24R^{Leu18Phe} virus transcript as analyzed by 2-tailed T-test.

APPENDIX C

EVOLUTIONARY ANALYSIS OF THE INTRACELLULAR DNA SENSOR IFI16

Introduction

The interactions between viruses and host innate immune receptors can determine the outcome of infection. Host pattern recognition receptors detect conserved pathogen-associated molecular patterns to initiate signaling pathways in the cell and induce an immune response (reviewed in Brubaker et al., 2015). These include receptors that detect foreign nucleic acids, which are a common feature of viral infections (Goubau et al., 2013; Sparrer and Gack, 2015). Viruses have evolved a number of mechanisms to inhibit these host nucleic acid sensors (Chan and Gack, 2016), highlighting critical interfaces between viruses and cells during the initial stage of infection. Since the outcome of these interactions can result in effective immune control or successful infection, there is strong selective pressure on both sides to maintain the upper hand. These ongoing conflicts thus create an evolutionary arms race between host nucleic acid sensors and viral antagonists. Studying these interfaces could elucidate viral immune-evasion strategies and the ways in which repeated antagonism has shaped the evolution of host immune responses.

Many immune factors are rapidly evolving, likely in response to recurrent episodes of selection imposed by pathogens, which act to fix changes that provide a

benefit to the host (Daugherty and Malik, 2012). Signatures of rapid evolution have been identified in some nucleic acid sensors, including the dsRNA sensors protein kinase R (PKR) (Elde et al., 2009; Rothenburg et al., 2009) and oligoadenylate synthetase 1 (OAS-1) (Hancks et al., 2015), and the cytoplasmic dsDNA sensor cyclic GMP-AMP synthase (cGAS) (Hancks et al., 2015). These evolutionary analyses have been used to identify key interfaces with viral factors, and the mechanisms by which host sensors have adapted to circumvent viral antagonists. This type of analysis could be applied to other host nucleic acid sensors to better understand the various ways viruses interact with the innate immune system.

Results

The host protein IFI16 is an intracellular DNA sensor belonging to the pyrin domain (PYD)-containing protein superfamily, and interacts with other pro-inflammatory factors to induce the production of interferons and interleukin-1β (IL-1β) (Orzalli et al., 2012; Unterholzner et al., 2010). IFI16 is a broadly antiviral sensor, capable of detecting DNA produced during infection with several different viruses, including vaccinia virus (VACV) (Unterholzner et al., 2010), herpes simplex virus-1 (HSV-1) (Orzalli et al., 2012), Kaposi's sarcoma-associated herpesvirus (KSHV) (Kerur et al., 2011), Epstein-Barr virus (EBV) (Ansari et al., 2013), and human immunodeficiency virus 1 (HIV-1) (Jakobsen et al., 2013). This broad specificity suggests that IFI16 is a critical immune defense protein and could be involved in a molecular arms race with viral antagonists.

I performed an evolutionary analysis to determine whether IFI16 is rapidly evolving and potentially engaged in a molecular arms race with viral pathogens. I

sequenced the IFI16 gene from a panel of 26 primate cell lines, and using two independent maximum-likelihood analyses, found evidence for strong positive selection in IFI16 ($p < 0.0005$). Branch-specific analyses indicated that ancient and recent episodes of selection have contributed to the adaptation of IFI16 throughout primate evolution, including events leading to the human lineage (Figure C.1). Further analyses of site-specific positive selection revealed a number of rapidly evolving amino acid positions throughout the IFI16 protein (Figure C.1). Patterns of selection dispersed across the protein are consistent with a long history of direct interaction with multiple antagonists (Daugherty and Malik, 2012). Our results are consistent with ancient and perhaps contemporary antagonists driving the rapid evolution of IFI16 in primates.

Discussion and future directions

The role for IFI16 during various viral infections and the signatures of positive selection I identified suggest that viruses likely possess mechanisms to antagonize IFI16. Viral proteins are known to interact with other pyrin domain (PYD)-containing proteins, including a poxvirus protein that interacts with the ASC-1 component of the inflammasome to inhibit apoptosis and inflammation (Johnston et al., 2005). Interactions between pyrin domains suggest that this could be a similar mechanism of action for virus proteins to contact IFI16. Indeed, recently, the human cytomegalovirus (HCMV) protein pUL83 was shown to interact with the IFI16 pyrin domain, preventing subsequent activation of immune signals (Li et al., 2013). Another potential IFI16 inhibitor is the HSV-1 protein ICP0, which induces the degradation of IFI16 in certain cell types (Cuchet-Lourenco et al., 2013; Orzalli et al., 2012; 2016). Studies to determine whether

residues in IFI16 under positive selection affect inhibition by either of these viral antagonists could reveal new insights into structural components of IFI16 important for such interactions. Changes at these interfaces could elicit fundamental consequences for biological function that determine the outcome of infection. Examining these interactions could help expose the complex mechanisms driving the evolution of host immune responses and viral strategies for immune evasion.

Materials and methods

Sequence analysis

Total RNA was extracted from primate fibroblast cell lines (Coriell), and cDNA was synthesized using a Maxima cDNA synthesis kit (Thermo Fisher). IFI16 sequences were PCR amplified from cDNA using primers in the 5' and 3' UTR of the human sequence, and DNA sequences obtained by Sanger sequencing.

Evolutionary analysis

DNA sequences were aligned using MUSCLE in default mode, and manually trimmed to remove indels. The alignment and a species tree representing currently accepted primate phylogeny (Perelman et al., 2011) were used as input files for analysis with PAML (Yang, 2007) or HyPhy software on Datamonkey.org (Delport et al., 2010).

References

Ansari, M.A., Singh, V.V., Dutta, S., Veettil, M.V., Dutta, D., Chikoti, L., Lu, J., Everly, D., and Chandran, B. (2013). Constitutive interferon-inducible protein 16-inflammasome activation during Epstein-Barr virus latency I, II, and III in B and epithelial cells. J. Virol. *87*, 8606–8623.

Brubaker, S.W., Bonham, K.S., Zanoni, I., and Kagan, J.C. (2015). Innate immune pattern recognition: a cell biological perspective. Annu. Rev. Immunol. *33*, 257–290.

Chan, Y.K., and Gack, M.U. (2016). Viral evasion of intracellular DNA and RNA sensing. Nat. Rev. Microbiol. *14*, 360–373.

Cuchet-Lourenco, D., Anderson, G., Sloan, E., Orr, A., and Everett, R.D. (2013). The viral ubiquitin ligase ICP0 is neither sufficient nor necessary for degradation of the cellular DNA sensor IFI16 during herpes simplex virus 1 infection. J. Virol. *87*, 13422-13432.

Daugherty, M.D., and Malik, H.S. (2012). Rules of engagement: molecular insights from host-virus arms races. Annu. Rev. Genet. *46*, 677–700.

Delport, W., Poon, A.F., Frost, S.D., and Kosakovsky Pond, S.L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics *26*, 2455–2457.

Elde, N.C., Child, S.J., Geballe, A.P., and Malik, H.S. (2009). Protein kinase R reveals an evolutionary model for defeating viral mimicry. Nature *457*, 485–489.

Goubau, D., Deddouche, S., and Reis e Sousa, C. (2013). Cytosolic sensing of viruses. Immunity *38*, 855–869.

Hancks, D.C., Hartley, M.K., Hagan, C., Clark, N.L., and Elde, N.C. (2015). Overlapping patterns of rapid evolution in the nucleic acid sensors cGAS and OAS1 suggest a common mechanism of pathogen antagonism and escape. PLoS Genet. *11*, e1005203.

Jakobsen, M.R., Bak, R.O., Andersen, A., Berg, R.K., Jensen, S.B., Tengchuan, J., Laustsen, A., Hansen, K., Ostergaard, L., Fitzgerald, K.A., Xiao, T.S., Mikkelsen, J.G., Mogensen, T.H., and Paludan, S.R. (2013). IFI16 senses DNA forms of the lentiviral replication cycle and controls HIV-1 replication. Proc. Natl. Acad. Sci. U.S.A. *110*, E4571-E4580.

Johnston, J.B., Barrett, J.W., Nazarian, S.H., Goodwin, M., Ricciuto, D., Wang, G., and McFadden, G. (2005). A poxvirus-encoded pyrin domain protein interacts with ASC-1 to inhibit host inflammatory and apoptotic responses to infection. Immunity *23*, 587–598.

Kerur, N., Veettil, M.V., Sharma-Walia, N., Bottero, V., Sadagopan, S., Otageri, P., and Chandran, B. (2011). IFI16 acts as a nuclear pathogen sensor to induce the inflammasome in response to Kaposi Sarcoma-associated herpesvirus infection. Cell Host Microbe *9*, 363–375.

Li, T., Chen, J., and Cristea, I.M. (2013). Human cytomegalovirus tegument protein pUL83 inhibits IFI16-mediated DNA sensing for immune evasion. Cell Host Microbe *14*, 591–599.

Orzalli, M.H., Broekema, N.M., and Knipe, D.M. (2016). Relative contributions of

herpes simplex virus 1 ICP0 and vhs to closs of cellular IFI16 vary in different human cell types. J. Virol. *90*, 8351–8359.

Orzalli, M.H., Deluca, N.A., and Knipe, D.M. (2012). Nuclear IFI16 induction of IRF-3 signaling during herpesviral infection and degradation of IFI16 by the viral ICP0 protein. Proc. Natl. Acad. Sci. U.S.A. *109*, E3008–E3017.

Perelman, P., Johnson, W.E., Roos, C., Seuánez, H.N., Horvath, J.E., Moreira, M.A., Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., Schneider M.P., Silva, A., O'Brien S.J., and Pecon-Slattery, J. (2011). A molecular phylogeny of living primates. PLoS Genet. *7*, e1001342.

Rothenburg, S., Seo, E.J., Gibbs, J.S., Dever, T.E., and Dittmar, K. (2009). Rapid evolution of protein kinase PKR alters sensitivity to viral inhibitors. Nat. Struct. Mol. Biol. *16*, 63–70.

Sparrer, K.M., and Gack, M.U. (2015). Intracellular detection of viral nucleic acids. Curr. Opin. Microbiol. *26*, 1–9.

Unterholzner, L., Keating, S.E., Baran, M., Horan, K.A., Jensen, S.B., Sharma, S., Sirois, C.M., Jin, T., Latz, E., Xiao, T.S., Fitzgerald, K.A., Paludan S.R., and Bowie, A.G. (2010). IFI16 is an innate immune sensor for intracellular DNA. Nat. Immunol. *11*, 997–1004.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.

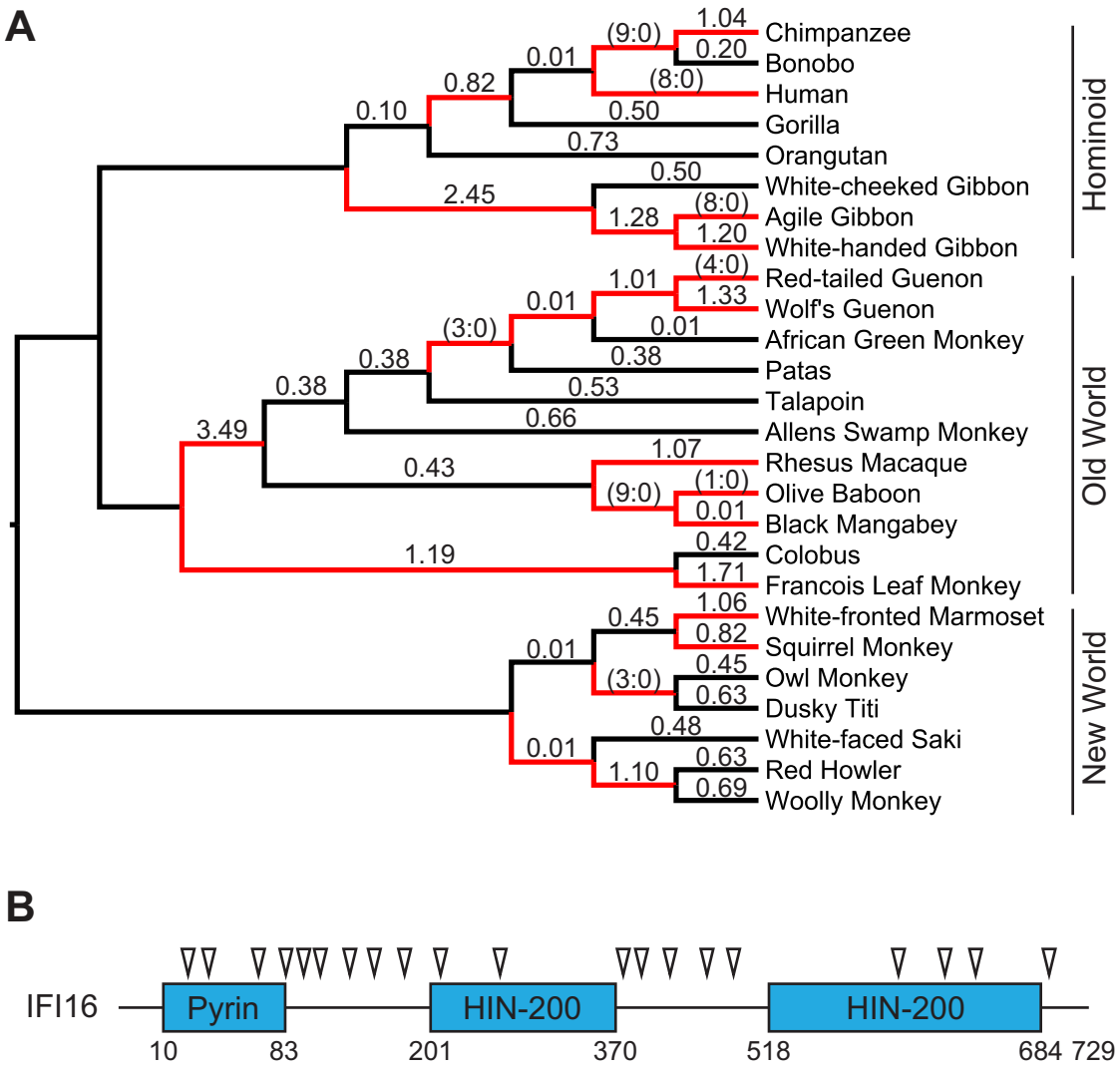


Figure C.1. Evolutionary analysis of IFI16 in primates. (A) Primate species tree with ω > 1 values from GA-BRANCH (HyPhy) shown in red, and dN/dS or nonsynonymous:0 synonymous amino acid substitutions from PAML analysis shown above each branch. (B) Schematic of the major domains of IFI16, with positively selected residues from PAML analysis (p > 0.95) indicated with open arrowheads above. Amino acid positions are shown below.